

Artificial neural network algorithm for analysis of Rutherford backscattering data

N. P. Barradas*

*Instituto Tecnológico e Nuclear, Reactor, Estrada Nacional 10, 2686-953 Sacavém, Portugal
and Centro de Física Nuclear da Universidade de Lisboa, Avenida Prof. Gama Pinto 2, 1699 Lisboa Codex, Portugal*

A. Vieira

*Instituto Tecnológico e Nuclear, Reactor, Estrada Nacional 10, 2686-953 Sacavém, Portugal
and Universidade Lusófona, Campo Grande 376, Lisboa, Portugal*

(Received 5 May 2000)

Rutherford backscattering (RBS) is a nondestructive, fully quantitative technique for accurately determining the compositional depth profile of thin films. The inverse RBS problem, which is to determine from the data the corresponding sample structure, is, however, in general ill posed. Skilled analysts use their knowledge and experience to recognize recurring features in the data and relate them to features in the sample structure. This is then followed by a detailed quantitative analysis. We have developed an artificial neural network (ANN) for the same purpose, applied to the specific case of Ge-implanted Si. The ANN was trained with thousands of constructed spectra of samples for which the structure is known. It thus learns how to interpret the spectrum of a given sample, without any knowledge of the physics involved. The ANN was then applied to experimental data from samples of unknown structure. The quantitative results obtained were compared with those given by traditional analysis methods and are excellent. The major advantage of ANNs over those other methods is that, after the time-consuming training phase, the analysis is instantaneous, which opens the door to automated on-line data analysis. Furthermore, the ANN was able to distinguish two different classes of data which are experimentally difficult to analyze. This opens the door to automated on-line optimization of the experimental conditions.

PACS number(s): 07.05.Mh, 82.80.Yc, 07.05.Kf, 68.55.Nq

I. INTRODUCTION

Rutherford backscattering (RBS), as well as other ion beam analysis (IBA) related techniques, is dedicated to the compositional analysis of samples in the few nm to the tens of μm range [1] and is extensively used in research laboratories. One of the reasons for its success is that it is fully quantitative, not needing to recur to certified samples (external standards), and the physics behind it is mostly classical.

The inverse RBS problem, which is to determine from the data the corresponding sample structure, is ill posed except in the most simple cases. While this is true of many inverse problems, skilled analysts have long sidetracked this barrier by using their experience-gained knowledge of what the RBS spectrum of a known sample looks like. This enables the analyst to recognize specific recurring features in RBS data and relate them to specific sample properties. This is then normally followed by a detailed quantitative analysis.

Until recently, analysis of RBS data (and other IBA techniques) had to be performed manually, with the help of interactive simulation codes, by highly skilled scientists. This time-consuming procedure precluded the analysis of large amounts of data. In particular, it prevented the development of an automated on-line analysis of data.

Recently, a code based on the simulated annealing (SA) algorithm [2,3] was presented which can analyze automatically RBS [4–6], elastic recoil detection (ERDA) [7], and

nonresonant nuclear reaction analysis (NRA) [8] spectra collected from a given sample. The analysis with SA is fast, taking the same order of magnitude as the time required to collect the data using a desktop PC, and only a small degree of user expertise is required to interpret the results.

However, it would nevertheless be highly desirable to have push-button instantaneous data analysis, particularly for specific systems of interest. This would enable batch analysis of samples, for instance, in quality control within a production line. Furthermore, the active participation of an expert during data collection is often required to optimize the experimental conditions. Automatic determination of optimal experimental conditions for a given sample is then another determining condition for successful automation of RBS experiments.

The aim of this paper is to present a code for the analysis of RBS data based on a supervised feedforward artificial neural network (ANN) algorithm [9], which is a very flexible scheme, capable of approximating an arbitrary unknown function. It is particularly well suited to high-dimensional nonlinear regression analysis with noisy signals and incomplete data. It was applied with success to many different fields, including infrared spectroscopy [10], the traveling salesman problem [11], and medicine [12].

Techniques such as SA, Bayesian inference and maximum entropy with the Markov chain Monte Carlo algorithm, and genetic algorithms have been recently applied to different experimental techniques such as RBS, particle-induced x-ray emission [13], ellipsometry [14], or x-ray fluorescence [15]. One of the main advantages of ANNs over those techniques is that they require some time (which in some cases

*Author to whom correspondence should be addressed. Electronic address: nunoni@itn1.itn.pt

can be very long) for the analysis of each data set, while ANNs, after the time-consuming initial training process, analyze the data instantaneously.

Another foremost ability of ANNs is to recognize recurring patterns in the input data. This is done without specific knowledge of the causes of the pattern. ANNs are then an ideal candidate to do automatically what RBS analysts have long done, which is to relate specific features of the data to specific properties of the sample. To test the capability of ANNs to perform this task, we chose a simple system, Ge implanted in Si [16], which has, however, many possible technological applications, in particular in the field of optoelectronics, with the potential for monolithic integration with Si technology [17].

II. RUTHERFORD BACKSCATTERING

A. Theory

In 1909 Geiger and Marsden observed what they called ‘‘diffuse scattering’’—that is, backward scattering—of MeV alpha particles by thin metal foils [18] (subsequently known as the ‘‘Rutherford experiment’’) and the following year measured the most probable angle of scattering of transmitted particles [19]. Based on these data, Rutherford developed his model of the atom and calculated the one-body scattering cross section using a Coulomb potential [20]. His predictions on the angular dependence of the cross section were confirmed by Geiger and Marsden [21] in 1913. Two-body calculations lead to correcting terms to Rutherford’s expression, and the final result, the so-called Rutherford cross section σ_{Ruth} , is

$$\sigma_{\text{Ruth}} = \left(\frac{Z_m Z_M e^2}{4E} \right)^2 \frac{4}{\sin^4 \alpha_{\text{scatt}}} \times \frac{[[1 - (m/M)^2 \sin^2 \alpha_{\text{scatt}}]^{1/2} + \cos \alpha_{\text{scatt}}]^2}{[1 - (m/M)^2 \sin^2 \alpha_{\text{scatt}}]^{1/2}}, \quad (1)$$

where Z_m and Z_M are the atomic number of the incident and target particles, respectively, m and M their masses, α_{scatt} the angle of scattering, E the incident particle energy, and e the electron charge. The target particle is supposed initially at rest. Deviations from the Rutherford cross section exist, at low energies due to electron screening of the nuclear charge [22] and at high energies due to nuclear reactions between the two particles.

In the collision, the backscattered particles lose a fraction of their initial energy. The ratio between their energy before and after scattering, E_0 and E , respectively, is the so-called kinematic factor K :

$$K = E/E_0 = \left[\frac{(M^2 - m^2 \sin^2 \alpha_{\text{scatt}})^{1/2} + m \cos \alpha_{\text{scatt}}}{M + m} \right]^2. \quad (2)$$

The scattering does not necessarily occur at the surface of a sample. As the particles of the incident beam penetrate the sample, they lose an amount of energy ΔE_{in} , mainly through interactions with the sample electrons. In the scattering, the particles lose energy (ΔE_K) according to Eq. (2). After scat-

tering, they further lose energy (ΔE_{out}) on their way out. The energy of the particle after being scattered at depth t and leaving the sample is then

$$E_{\text{det}} = E_0 - \Delta E_{\text{in}} - \Delta E_K - \Delta E_{\text{out}}, \quad (3)$$

where

$$\Delta E_{\text{in}} = \int_0^t S[E(x)] dx, \quad (4)$$

$$\Delta E_K = (1 - K)(E_0 - \Delta E_{\text{in}}), \quad (5)$$

$$\Delta E_{\text{out}} = - \int_{t/\cos \alpha_{\text{scatt}}}^0 S[E(x)] dx, \quad (6)$$

and

$$S[E(x)] = \frac{dE}{dx} \quad (7)$$

is the energy loss per unit path length, which is a function of the energy of the particle, which in its turn depends on the path already traversed. These energy loss values have been determined experimentally for many incident beams and targets, and are available in tabulated form [23]. Note that Eqs. (4) and (6) are valid only for normal incidence. When the sample is tilted at an angle θ_{inc} with the normal to the sample, the integral limits are scaled according to $\cos \theta_{\text{inc}}$, in order to account for the longer path of the beam through the sample.

Equations (1)–(3) constitute the basis of the Rutherford backscattering technique. The energy of particles backscattered from the surface of a sample depends only on known parameters and on the mass of the scattering center. Hence, from Eq. (2), one can determine *what* the sample is made of. The Rutherford cross section depends also only on known parameters (after the target particles have been identified); hence, from Eq. (1) one can determine *how much* of a given element the sample contains. Finally, from the energy loss and Eq. (3) one can determine *where* a certain element is located in a given sample. This is quantitative in an absolute way: that is, it does not depend on external standards, which is the main factor making RBS (and other ion beam analysis techniques) attractive. A detailed description of the numerical generation of RBS spectra can be found in Ref. [24].

Equation (3) provides a depth scale, i.e., a conversion between detected energy and depth; however, it is different for each target element and is also not linear. For instance, the signal due to scattering from a heavy element located deep in a sample can be, and often is, superimposed onto the signal due to a lighter element located near the surface of the sample. This renders analysis difficult. While it is very easy to calculate the expected spectrum from a given beam and sample (and parameters of the detection system) by discretizing the sample into many sublayers, the inverse problem of calculating the sample structure from a measured spectrum is considerably more difficult.

Traditionally, analysis for all but the most simple cases has been done with the aid of computer programs. These are almost all interactive, requiring the analyst to guess the

structure of the sample, calculate a spectrum, compare it with the data, and iterate until a sufficiently good match is obtained. A recent review of available codes is given in Ref. [25]. Recently, a code NDF, based on the SA algorithm [2,3], has been presented which effectively solved the inverse RBS problem [4].

The theoretical model presented above has some limitations. First, it considers only a single large-angle scattering event. It disregards the existence of plural scattering, in which several large-angle scattering events take place, and multiple scattering, in which many small-angle scattering events take place before or after the backscattering occurs. These effects lead to an enhanced yield at low energies, that is, a low-energy background (in some cases, particularly low-energy heavy ions, also a higher-energy background can occur [26]), as well as to degradation of the energy resolution [27,28].

Second, in this model we consider only the energy resolution of the setup. We have disregarded energy straggling due to the statistical nature of the energy loss process, to the energy and angular spread of the analyzing beam, geometric straggling caused by the finite size of the beam spot and detector, and multiple scattering. The influence of all these factors can be calculated [29,30], but it is rather involved and can be time-consuming.

B. Experimental details

Ge was implanted into Si with nominal beam fluence (also called implanted dose) between 10 and 25×10^{15} Ge atoms/cm², to projected ranges between 491 and 776 Å, that is, between 245 and 338×10^{15} Si atoms/cm². The RBS analysis utilized a 1.5 -MeV He⁺ beam. Both normal incidence and a 7° misalignment were used. Two detectors were used in the IBM geometry (in which the incident beam, the detected beam, and the normal to the sample all lie in the same horizontal plane), with scattering angles 165° and 133° . The product of detector solid angle with analyzing beam fluence was between 4.3 and 124.2 $\mu\text{C msr}$.

III. ARTIFICIAL NEURAL NETWORKS

A. Theory

An artificial neural network is a flexible scheme capable of approximating any arbitrary unknown function [9]. It is particularly well suited to high-dimensional nonlinear regression analysis with noisy signals and incomplete data [31]. A feedforward neural network consists of an array of input nodes connected to an array of output nodes through the nodes of successive intermediate layers. Each connection between nodes has a weight, initially random, which is adjusted during a training process. The output of each node of a specific layer is a function of the sum on the weighted signals coming from the previous layer. The crucial points in the construction of an ANN are the selection of inputs and outputs, the architecture of the ANN, that is, the number of layers and of nodes in each layer, and finally, the training algorithm.

In supervised learning [9], the training is done by presenting a large set of examples, called the training set, to the network. Each example consists of a set of inputs presented

to the input layer and the respective set of desired outputs presented to the output layer. The training set normally consists of experimental data. However, when a good computational model is available, the training set may be constructed with simulated experimental data.

The training of an ANN is in general very time-consuming. After being trained, however, the experimental data are analyzed almost instantaneously. Furthermore, care must be taken in order to use an adequate training set, representative of all possible experimental situations. In many cases this is not feasible, and the sampling space must be restricted to a specific subdomain. This means that ANNs are best applied to specific, well-defined problems.

There are several training algorithms for supervised learning, the backpropagation algorithm being the most widely used. In this algorithm the training is performed by minimizing the sum of square errors function over the N elements of the training set:

$$E = \frac{1}{2} \sum_{n=1}^N (y_n - o_n)^2. \quad (8)$$

In this equation y is the output given by the network (for simplicity, we consider a single output) and o_i is the desired output. By minimizing this equation with respect to the weights w_{ij}^k connecting the node i of the last hidden layer k to nodes j of the output layer, we obtain the following recursive relation for updating the weights after each iteration [32]:

$$w_{ij}^k(t+1) = w_{ij}^k(t) + \eta \delta^k x_i + \alpha [w_{ij}^k(t) - w_{ij}^k(t-1)], \quad (9)$$

where x_i is the output of the node i in layer k and δ^k is given by

$$\delta^k = y(1-y)(o-y). \quad (10)$$

To actualize the weights of a given layer $m-1$, the error is backpropagated substituting δ^k by

$$\delta_i^{m-1} = x_i(1-x_i) \sum_j \delta_j^m w_{ij}^{m-1}, \quad (11)$$

where x_i is the output of node i in layer $m-1$, and for the last hidden layer k , δ_j^k is defined as δ^k . The learning rate parameter η is usually set below 0.2 . A momentum term α is added to avoid trapping in local minima. In our case we set $\alpha = 0.2$. The output of each node is taken to be the sigmoid function $f(x)$ of the weighted sum x of the outputs from all the nodes of the previous layer:

$$f(x) = (1 + e^{-(x-\theta)})^{-1}. \quad (12)$$

The parameter θ is a threshold that usually is set to zero. At each iteration the results provided by the ANN are compared with a sample of examples not used in the training, called the test set. The figure of merit of the network is the final mean-square error (MSE), given by

$$\varepsilon_{\text{MSE}} = \sum_j \frac{(y_i - o_i)^2}{(o_i - \bar{o})^2}. \quad (13)$$

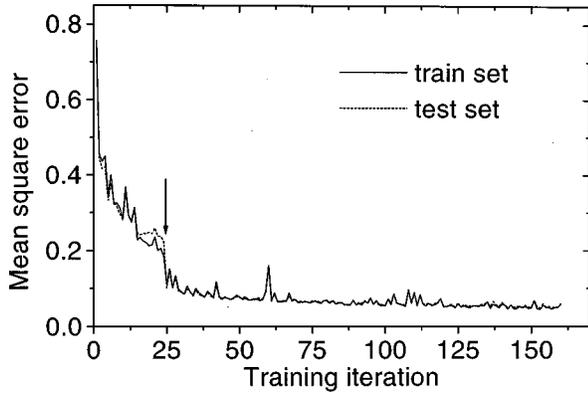


FIG. 1. Mean-square error as a function of the number of training iterations for the train set and the test set. The arrow marks the iteration where the error reached 20% and all cases with error larger than 40% were excluded from the train and test sets.

The sum is over the training set, and \bar{o} is the mean value of the output used in the training data. The training process is stopped after some convergence criterion is reached.

During training, the MSE of the training set decreases with the number of iterations, while for the test set it reaches a minimum, after which it starts to increase. This occurs due to overfitting of the net to the training data, thus decreasing its generalization capabilities. To overcome this problem we have to define a stop criterion. The most widely used recipe is to stop the training process when the MSE of the test set is minimum (see Fig. 1). Another possibility to avoid overfitting is the addition of noise to the training data. In our case both techniques are simultaneously used.

B. Network architecture

The selection of an adequate network architecture is one of the most important aspects to consider in using ANNs. Although an ANN with a single hidden layer is in principle capable of classifying any pattern (provided that an infinite number of training examples is available), in many cases it is useful to use more intermediate layers. This can have several advantages. For example, by introducing intermediate layers with a number of nodes smaller than the number of input nodes, the dimension of the input space is reduced through a projection of the data into a lower-dimensional space. In practice, this is similar to applying feature extraction methods on the input data, like principal component analysis [33], which is a useful technique when a large number of highly correlated inputs exist. This leads to a more robust map between the inputs and outputs, and with a lower error. Since the sampling space is effectively reduced, dimensionality reduction is also appropriated in cases where only a few training examples are available.

On the other hand, using a large number of intermediate layers may cause some difficulties. Due to the increase of optimization parameters (i.e., the connection weights) the risk of the ANN becoming trapped in a local minimum is higher. The risk of overfitting is also higher, and, finally, it takes more time to train ANNs with a large number of layers.

Several techniques of alleviating these difficulties exist. One of them is weight decay [34], which essentially eliminates less significant nodes or weights from the network. A

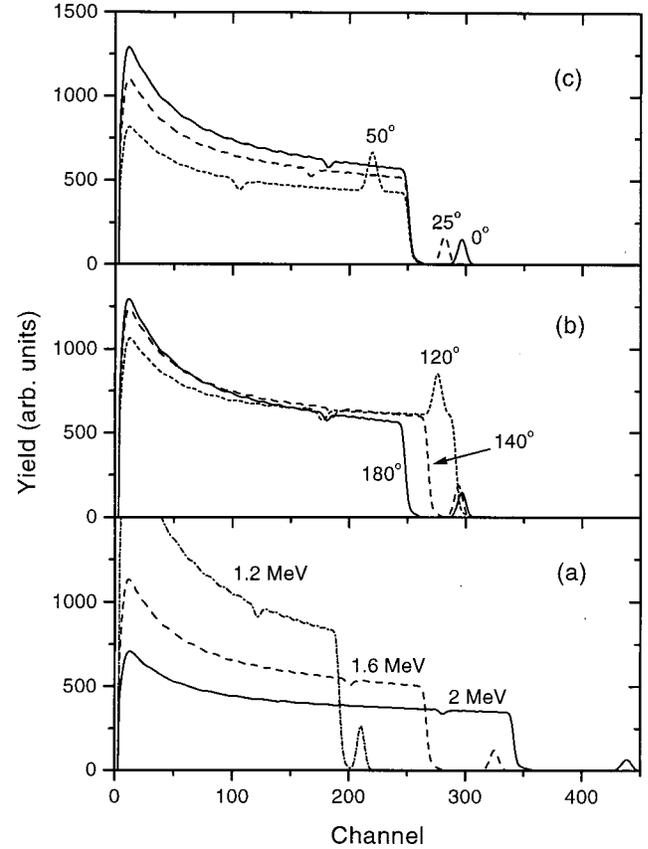


FIG. 2. Spectra calculated for different experimental conditions for a 25-Å-thick Ge δ layer located under a 400-nm-thick Si layer: (a) Beam energy $E_0=1.2, 1.6,$ and 2 MeV. (b) Scattering angle $\alpha_{\text{scatt}}=120^\circ, 140^\circ,$ and 180° . (c) Angle of incidence $\theta_{\text{inc}}=0^\circ$ (normal incidence), $25^\circ,$ and 50° .

different technique consists in using node growth and pruning algorithms in which new nodes and layers are added or deleted based on the sensitivity of the output to their weights [9]. Both these methods have the disadvantage of demanding the computation, or some approximation, of the second derivative of the error function with respect to all the networks weights, i.e., the Hessian matrix, which is a time-consuming task.

In principle, the determination of determining the best network configuration, i.e., the network with the lowest mean-square error, is an optimization problem that can be solved by standard optimization algorithms, such as simulated annealing or genetic algorithms [35]. These have the major drawback that they require the training of thousands of ANNs and the evaluation of their performance.

C. Application to Rutherford backscattering

We present in Fig. 2 spectra calculated for different experimental conditions, for a Ge sample 25 Å thick (i.e., 11.1×10^{15} Ge atoms/cm²) δ layer located under 400 nm of Si (i.e., 2×10^{18} Si atoms/cm²). It is clear that, for one given sample, very different spectra can be obtained. Some of the possible spectra are relatively simple to analyze (large and

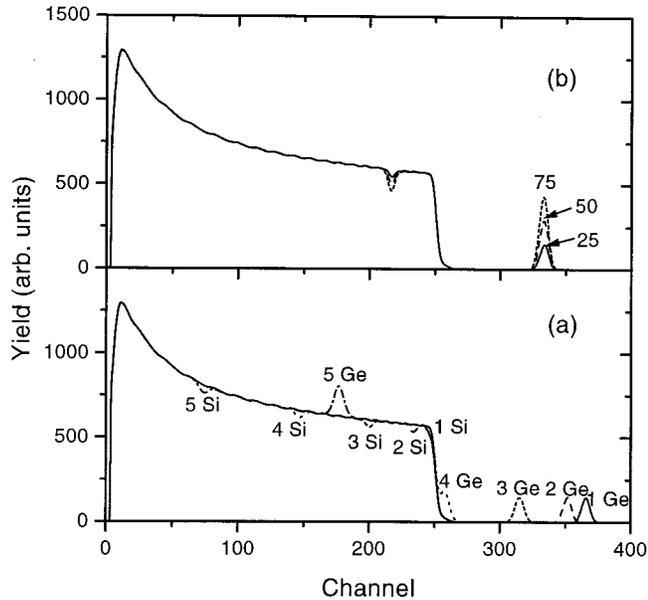


FIG. 3. Calculated spectra for a Ge δ layer in Si: (a) 10 \AA thick, located at different depths: 1–20, 2–100, 3–300, 4–600, and 5–1000 nm. The position of the Ge peaks and of the corresponding Si dips is marked. (b) 25, 50, and 75 \AA thick, located under a 200-nm-thick Si layer.

well-separated signals) and others relatively complicated (e.g., when the signals of Ge and Si are superimposed). The purpose of an ANN dedicated to the analysis of RBS spectra, and in particular to the analysis of Ge in Si, is, first, to determine the amount of Ge present and the depth at which it is located. This should be achieved independently of the experimental conditions, that is, with any of the spectra shown in Fig. 2 given as input, and as long as the experimental conditions are also part of the input, the ANN should be able to determine that the amount of Ge is $11.1 \times 10^{15} \text{ Ge/cm}^2$ and that its central point is located at a $2005.5 \times 10^{15} \text{ atom/cm}^2$ depth. Second, if this is not possible due to data that are too hard to analyze (e.g., with partially or completely superimposed Ge and Si signals), the ANN should be able to recognize that fact and return feedback with suggested optimized experimental conditions. Eventually, one could train different networks, specialized for different classes of data (e.g., separated signals versus superimposed signals).

We show in Fig. 3(b) the spectra expected for a 25- \AA -thick Ge δ layer located at different depths in Si, for a 1.5-MeV He^+ beam at normal incidence and detected at 165° scattering angle. Given the experimental conditions, the ANN can use the position of the Ge signal relative to the Si signal to determine the Ge depth. The dip in the Si signal could in principle also be used, but it is small and thus easily masked by statistical noise. Furthermore, given the experimental conditions, the ANN can use the size of the Ge signal to determine the amount of Ge as shown in Fig. 3(b). Again, the size of the Si dip provides only limited information. On the other hand, the total yield in the Ge signal is directly proportional to the product of detector solid angle and analyzing beam fluence, which must also be an input. However, it should be noted that the ANN can use the size of the Si signal, which is also proportional to the product of detector

solid angle and analyzing beam fluence, as a normalizing factor to help determine the Ge dose.

The data must be preprocessed before being presented to the ANN. The most important step is to normalize all the inputs and outputs. Since the signals have a wide variation range (three orders of magnitude), we chose to apply the following transformations: for the output depth $y = \sqrt{y+b}$, and for the implanted dose $y = \ln(y+b)$. The threshold value was set to $b=5$ and was included in order not to give extreme importance to cases with very low values. For the depth we prefer to use the square root since its variation within the training set is lower than for the dose.

The yield is first normalized to the charge–solid-angle product and then normalized to values between 0 and 1 (1 corresponding to the largest single-channel yield observed in the training and test sets). The total yield (signal area) is also used as an input. The other inputs are the beam energy and energy resolution, the angle of incidence and the scattering angle, and the input data consisting of 128 channels.

IV. ANN IN RBS ANALYSIS OF Ge IMPLANTS INTO Si

A. Construction of the training set

The training set must cover a wide range of realistic experimental conditions and of possible Ge doses and depths. We generated a training set consisting of theoretical spectra simulating different Ge implants into Si, to doses between 10^{14} and $10^{18} \text{ Ge atoms/cm}^2$ and depths between 1 and 1500 nm (which corresponds to an implant energy of about 2.8 MeV). Each implant was simulated as a perfect Gaussian, with width given by the code TRIM [23] for the corresponding implantation depth subject to an up to 20% random change. The spectra were calculated for different beam and detection parameters chosen at random, in order to simulate a very broad range of realistic experimental conditions. The beam was He, with energy between 1 and 2 MeV and resolution between 13 and 40 keV full width at half maximum (FWHM). The scattering angle was between 130° and 180° , and the detection angle was between -30° and 30° , considering the normal to the sample in the plane defined by the beam (IBM geometry). The collected charge (beam fluence) was between 0.2 and 250 μC for a solid angle of 1 msr. We then added the contribution of pulse pileup to the theoretical spectra [36], and finally we added Poisson noise in order to simulate experimental data as closely as possible. We did not include the effect of plural and multiple scattering, which lead to a low-energy background. The reason was that they can only be calculated by extremely time-consuming Monte Carlo methods [28], and furthermore the theoretical discussion on the origin of the low-energy tails observed experimentally is still ongoing [37].

The space of implant, beam, and detection parameters (training space) utilized was not uniform. Instead, more training spectra were generated for the beam and detection conditions that are more common in real experimental situations. We thus provide a larger amount of training examples corresponding to common and to difficult cases. The training space utilized is shown in Fig. 4.

For the implant conditions, we chose an almost uniform (logarithmically) distribution for the Ge dose, decreasing slowly for extremely high doses that only seldom are im-

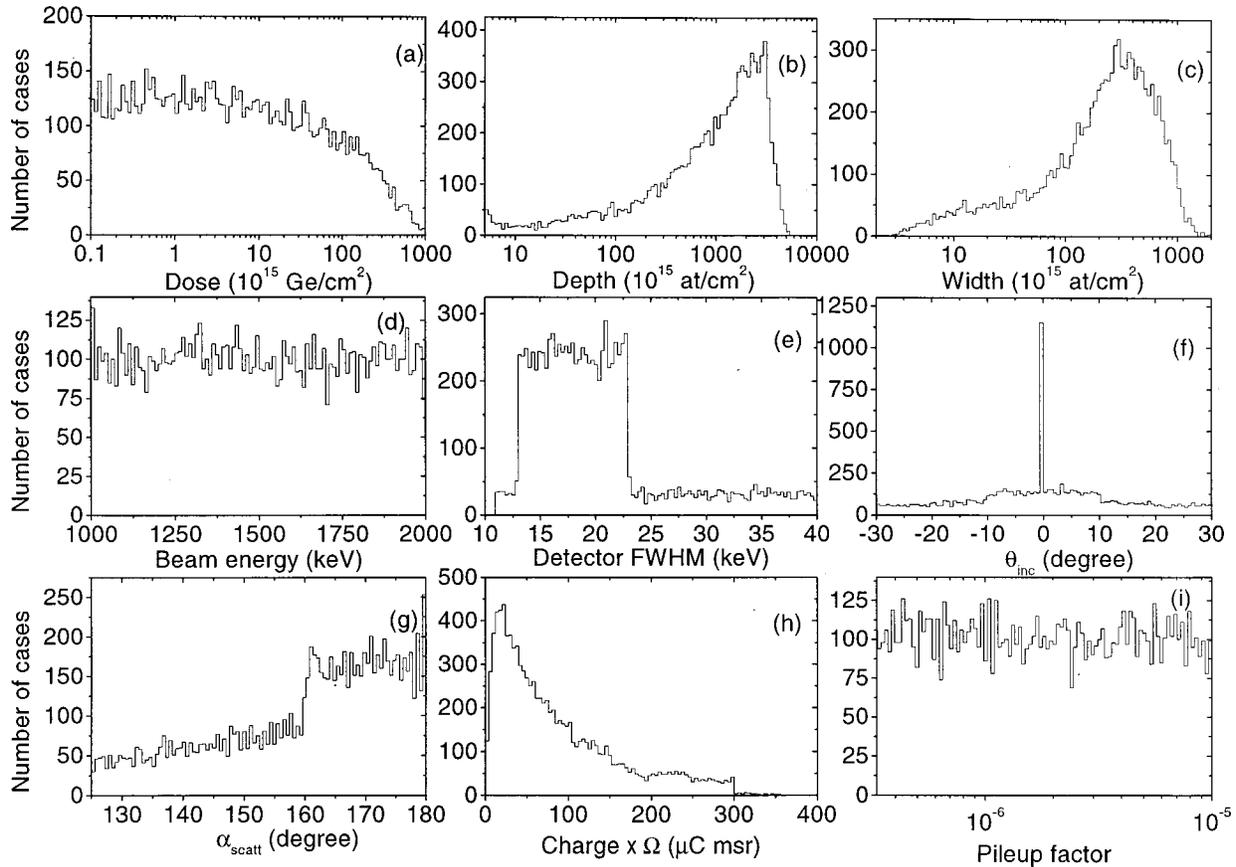


FIG. 4. Space of sample, beam, and detection conditions utilized to generate the training and test sets.

planted (due to cost and time considerations). A larger weight was given to larger implant depths, since for these cases the analysis is more difficult due to the superposition of the Ge and Si signals. As for the width of the implant, it is strictly correlated to the implant depth as mentioned above and as is shown in Fig. 5(a).

The distribution of the beam energy was uniform, as 1–2 MeV are common energies used for He^+ RBS, and different laboratories use different values. The distribution of the detector resolution is given by the superposition of two different uniform distributions, the first one from 11 to 40 keV (respectively, very good and very poor values) and the second one, more probable, from 13 to 22 keV, which are normal experimental values. For the angle of incidence (relative to the normal of the sample), we chose a superposition of three distributions: a broad one from -30° to $+30^\circ$ to cover most of the experimental range used, a narrower one from -10° to $+10^\circ$ to cover the normal range (for instance, a tilt angle of about 7° is commonly used to avoid accidental channeling), and a δ distribution for $\theta=0^\circ$, as most of the experiments are done at normal incidence. For the angle of scattering we chose a uniform distribution between 160° and 180° , which is the range of most common values, superimposed to a lower probability distribution decreasing to 125° , covering the vast majority of cases. The distribution of the charge–solid-angle product peaks at $20 \mu\text{C msr}$ and then decreases slowly to very high values. Finally, the pileup factor, which determines the size of the pileup correction, was given by a logarithmically uniform distribution covering a broad range of values.

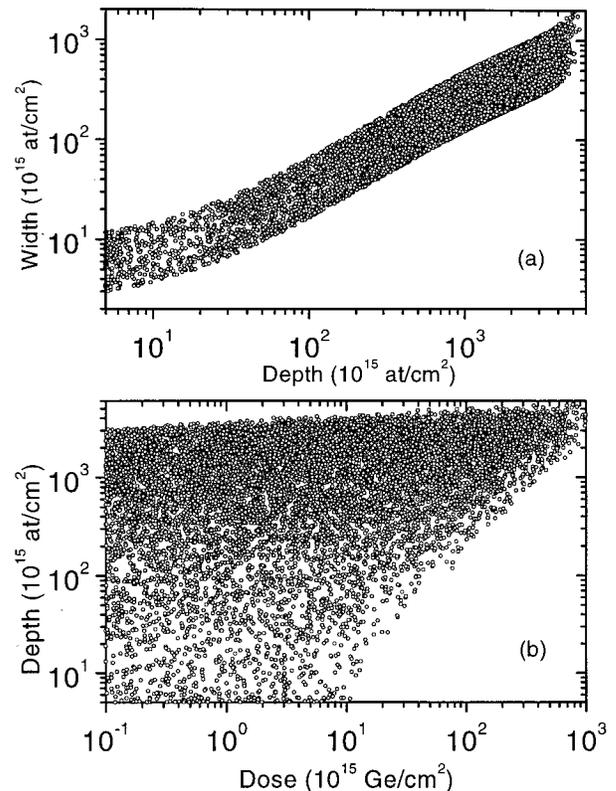


FIG. 5. Correlations in the training space, between (a) width and depth of implant and (b) depth and implanted dose. Each point is one case.

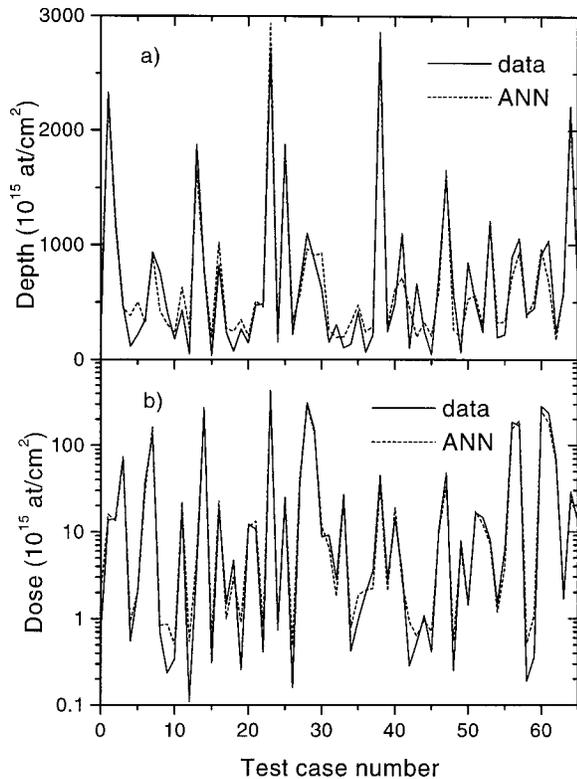


FIG. 6. Results for a series of the test set spectra that the ANN identified as being correctly analyzed. The original Ge dose and depth with which the spectra were generated are identified as data and the results given by the neural network as ANN.

Finally, we should note that only Gaussian distributions 99% inside the sample were generated. This means that shallow implants with high doses, where the implant shape deviates significantly from Gaussian, were not considered. These distributions with simultaneously high dose and small depth were hence not generated, as shown in Fig. 5(b).

B. Optimization of the architecture

As mentioned in Sec. III, the architecture of an ANN is a fundamental factor in determining its behavior and quality. The optimization methods mentioned above, such as simulated annealing or genetic algorithms, have the major drawback that they require the evaluation of the performance of thousands of ANNs, which is in general very time-consuming. In our case the training of each network with 10 000 examples takes about 10 h to be completed in a Pentium III computer running at 500 MHz. So we have to abandon a systematic approach to determine the best network configuration in favor of a more modest approach based on testing several different configurations.

We tested several ANNs, with between one and five hidden layers, and considering different numbers of nodes in each hidden layer. In all cases we trained the ANNs using the same training and test sets consisting of 4500 and 500 generated spectra, respectively. When the MSE in the train set reached 20%, we excluded from the training and test sets all the cases with error larger than 40% and continued training until a minimum on the test set MSE was reached (see Fig. 1). The results obtained are given in Table I.

TABLE I. Tested ANN architectures. I and O represent the input and output data, respectively. The intermediate numbers represent the number of nodes in each hidden layer. The $(I,100,80,50,O)$ ANN was chosen as the reference ANN.

Architecture	Train set error	Test set error
$(I,100,O)$	6.3	11.7
$(I,250,O)$	5.2	10.1
$(I,100,80,O)$	3.6	5.3
$(I,100,50,20,O)$	4.2	5.1
$(I,100,80,50,O)$	3.0	4.1
$(I,100,80,80,O)$	2.8	4.7
$(I,100,50,100,O)$	3.0	4.2
$(I,100,80,80,50,O)$	3.2	4.1
$(I,100,80,50,30,20,O)$	3.8	5.3

First of all, using a single hidden layer, the error in both the training and test sets decreases when the number of nodes increases from 100 to 250. However, the error remains very large, above 10% on the test set, which shows that a single hidden layer is inefficient classifying RBS spectra.

The addition of an extra hidden layer with 80 nodes to a previous 100-node layer has a beneficial effect by decreasing the MSE by a factor of 2. Increasing the number of layers to 3 but decreasing at the same time the number of nodes leads to a slightly worse performance. We hence kept the previous two-hidden-layer structure and added one extra layer with 50 nodes, which led to a further reduction in the test set error. This indicates that the inputs are correlated, but not too much.

Increasing further the number of nodes of the third hidden layer to 80 decreases the error in the training set, but the error in the test set becomes worse, in a clear sign that over-training occurred. In the same way, increasing the number of hidden layers to 4 or to 5 also leads to no reduction or even to an increase in the test set error. The error for these more complex architectures might decrease further using a larger number of spectra in the training set. However, this would be at the cost of a large increase in the time required to train the ANN.

We finally chose the architecture consisting of three hidden layers, with 100, 80, and 50 nodes. The results for some of the spectra are shown in Fig. 6. This is the ANN with the smallest test set error. One ANN tested has an equally small error on the test set, but with four hidden layers it is unnecessarily more complex and, furthermore, has a slightly worse error on the train set.

C. Reliability and robustness

It was mentioned above that, when the average error reached 20%, all the cases with error superior to 40% were excluded from the training process. The results for some of the spectra are shown in Fig. 6. This corresponded to about 13% of the cases in the training set, meaning that the ANN developed could not analyze a significant fraction of the data, posing the question of its reliability. We will now examine which cases were so eliminated. We show in Fig. 7 the depth versus dose diagram for the excluded cases. This compares with Fig. 5(b), where a similar diagram is shown, but for the entire initial training space. It is observed that the cases eliminated correspond to two broad classes: one in-

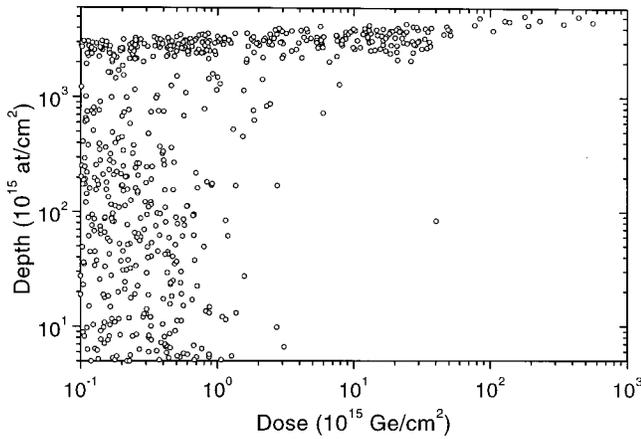


FIG. 7. Dose and depth of the spectra, with error larger than 40%, eliminated from the training set at a 20% average error level. Each point is one case eliminated.

cludes spectra with implanted depth larger than approximately 2000×10^{15} atom/cm², which we will call class *B*, and the other includes spectra with doses smaller than approximately 3×10^{15} atom/cm², which we will call class *C*. Class *A* corresponds to spectra with implanted depth smaller than 2000×10^{15} atom/cm² and dose larger than 3×10^{15} atom/cm², of which very few were eliminated.

We show in Fig. 8 the density of eliminated cases as a function of implanted depth and of implanted dose. From Fig. 8(a) it is concluded that samples that have both very small implant doses and large implant depths, that is, samples that belong simultaneously to classes *B* and *C*, have a much enhanced probability of having a large error and hence of being rejected at the end of the first training phase. Samples with higher implant doses are only rejected when the implant depth is very large, that is, if they belong to class *B*. On the other hand, we see from Fig. 8(b) that samples with smaller implant depths are only eliminated from the training process if they have very small implant doses, that is, if they belong to class *C*.

Some of the rejected spectra from classes *B* and *C* are shown in Fig. 9, where the reasons behind the rejection become apparent. Class *B* corresponds to spectra where the Ge is so deep in the sample that, taken the analyzing beam energy and the experimental conditions into account, the Ge signal becomes superimposed onto the Si signal. This leads to a spectral shape that is significantly different from the majority of spectra, in which the signals are well separated. It corresponds to a class of spectra that is difficult to analyze, even when traditional methods are used, also due to the signal superposition. The experimentalist's solution, when confronted with one such spectrum, is normally to try to change the experimental conditions in order to obtain separated signals. This can be done by increasing the beam energy or, if the sample was measured at a tilt angle θ_{inc} away from normal incidence, use $\theta_{inc} = 0^\circ$.

As for class *C*, it corresponds to spectra where the Ge signal is extremely small compared with the Si signal or with the pileup background. Without the logarithmic scale in Fig. 9(a), the Ge signal would go unnoticed. Furthermore, in real samples small levels of impurities in the Si signal could mask the Ge signal even further. Quantitative analysis, al-

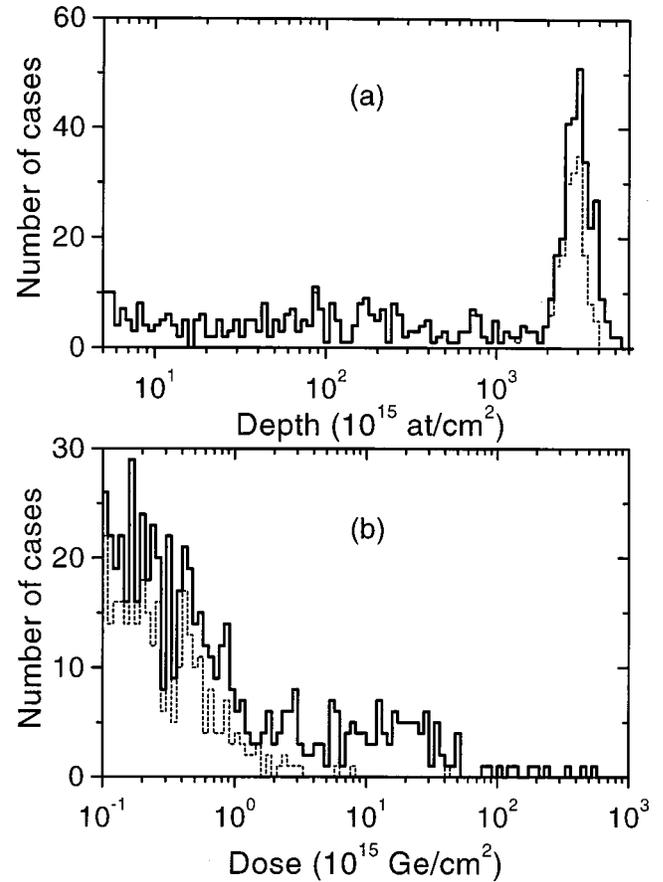


FIG. 8. Density of eliminated cases as a function of (a) implanted depth—the dashed line corresponds to cases with dose smaller than 3×10^{15} atom/cm²—and (b) implanted dose. The dashed line corresponds to cases with depth smaller than 2000×10^{15} atom/cm².

though still possible in the cases shown (and altogether hopeless for still smaller doses), becomes difficult and, in any case, leads to large errors. The experimentalist's solution for such cases is either to use a different technique with a better sensitivity to Ge or to increase the He beam fluence in order to obtain better counting statistics.

It should be noticed that the developed ANN effectively separated the input spectra into three distinct classes: that of the “normal,” or easier, cases to analyze and two different classes that correspond to samples that are experimentally difficult to analyze, for completely different reasons and requiring completely different actions from the experimentalist. This opens the door to on-line optimization of the experimental parameters, as, once feedback from the ANN to the controlling system of the accelerator is incorporated, automation of the optimization process can be achieved.

To test the sensitivity of the ANN to mistakes in the experimental data, we generated two different sets of 500 theoretical spectra each. All the parameters were in the same range as that used for the training set, but instead of simulating Ge implants, we simulated implants with the neighboring elements Ga and As. The mass resolution of RBS is given by the energy difference ΔE due to scattering from elements with mass around M and a mass difference of ΔM . It is, for a scattering angle of 180° ,

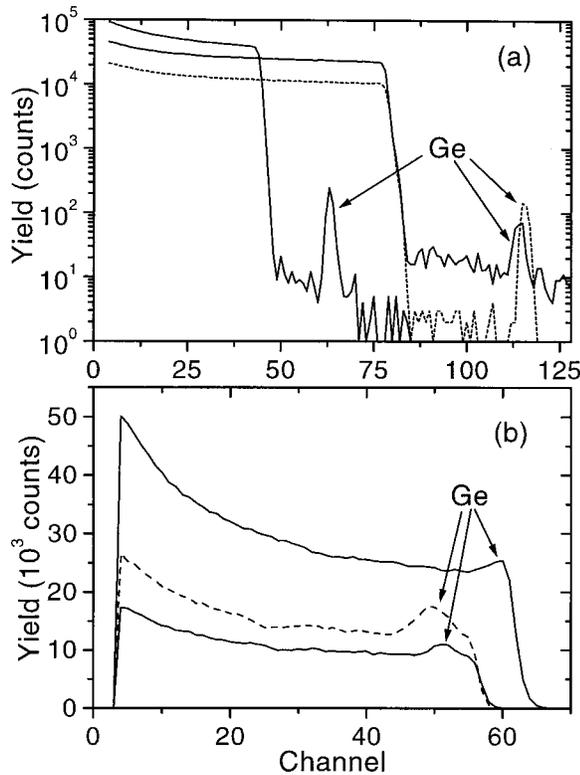


FIG. 9. (a) Spectra from class C. (b) Spectra from class B.

$$\Delta E = 4E_0(m/M^2)\Delta M, \quad (14)$$

where E_0 is the beam energy and m is the mass of the beam species. For 1.5-MeV⁴ He⁺ scattering off Ge, we have

$$\Delta E/\Delta M = 4.55 \text{ keV/amu}. \quad (15)$$

This leads to an energy difference of 13.1 and 10.5 keV between the Ga and Ge and As and Ge signals, respectively. These values are smaller than the energy resolution normally achieved with surface barrier detectors, since, as said above, the training set was generated with resolution between 13 and 40 keV. Furthermore, according to Eq. (1), the Rutherford cross section is similar for the three elements, differing 13% between Ga and As. This means that the difference is small enough that the ANN trained for Ge implants should still be able to analyze the Ga and As implant test data, albeit with a systematically larger error.

Utilizing the reference ANN, the error was 4.6% on the Ga set and 14.0% on the As sets. As expected, this is larger than the errors on the Ge training and test sets, which were 3.0% and 4.1% respectively. However, at first sight it could seem surprising that the error in the Ga set is smaller than that in the As set, because the mass difference between Ga and Ge is larger than that between As and Ge. We show in Fig. 10 generated spectra corresponding to an implanted dose of 25.9×10^{15} atom/cm² of Ge, Ga, and As. The Ga and Ge peaks are approximately the same shape and height. The main difference between them is their position, the Ga peak being located slightly to the left of the Ge peak due to the lower kinematic factor for Ga as given by Eq. (2). On the other hand, the height of the As peak is considerably larger than that of the Ge peak. The As peak is furthermore nar-

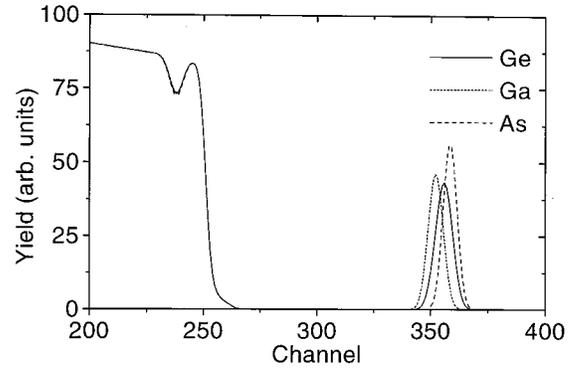


FIG. 10. Generated spectra for 25.9×10^{15} atom/cm² implants of Ge, Ga, and As.

rower than the Ge peak. It is this similarity between the Ga and Ge peaks and this difference between the As and Ge peaks that leads to only a modest increase in the error when the Ge-trained ANN is used on Ga data and to a large increase when it is used on As data.

The reason behind the shape of the Ga, Ge, and As peaks lies on their isotopic distribution. Natural Ga has two main isotopes ⁶⁹Ga and ⁷¹Ga with abundance 60.1 and 39.9, respectively. Although the scattering cross section is the same for the two isotopes, the kinematic factor is not, which leads to a slight broadening of the Ga peak. Natural Ge has three isotopes ⁷⁰Ge, ⁷²Ge, and ⁷⁴Ge with abundance 21.2, 27.7, and 35.9. On the one hand, this leads to an even larger broadening of the Ge peak than that of the Ga peak, which explains why the height of the Ge peak is slightly smaller than that of the Ga peak, even when its area is larger by a factor of $(Z_{\text{Ge}}/Z_{\text{Ga}})^2 = 1.065$. On the other hand, one of the Ge isotopes (⁷⁰Ge) is lighter than one of the Ga isotopes (⁷¹Ga), which leads to a larger degree of superposition between the Ga and Ge peaks than could be otherwise expected. On the contrary, As has a single isotope ⁷⁵As, leading to the narrow and large peak observed.

D. Comparison with experimental results

We applied the neural network, with the fixed weights obtained during the supervised training, to real data. We analyzed ten spectra collected in different experimental conditions. The total analysis time was less than 1 s. While during the supervised training the Ge dose and depth are given to the ANN, they are now outputs calculated directly from the experimental data. We compared the results obtained with the values determined using NDF [4,38]. The results are given in Table II and are excellent. It should be noted that the first 25 channels of each spectrum were disregarded, in order to eliminate the region where the plural scattering effect is larger.

While the precision achieved is worse than what can be obtained with the usual data fitting, simulation, or calculation methods, we stress that the analysis was fully automatic and performed practically instantaneously, hence being well suited for on-line analysis purposes. Furthermore, a fast local search algorithm could easily make the small improvements necessary to obtain precise values.

TABLE II. Experimental data results.

Sample	Values derived with NDF		Values obtained with the ANN	
	Ge (10^{15} atom/cm 2)	Depth (10^{15} atom/cm 2)	Ge (10^{15} atom/cm 2)	Depth (10^{15} atom/cm 2)
1	16.7	332.8	16.3	267.4
2	14.4	302.3	12.4	223.7
3	13.6	318.2	15.3	307.2
4	14.7	334.5	12.4	236.8
5	15.7	378.0	10.6	245.7
6	9.3	250.7	12.2	308.7
7	26.8	246.3	27.9	214.3
8	9.8	349.1	12.7	359.4
9	9.6	356.8	12.6	384.7
10	9.7	316.3	11.9	329.4

This is demonstrated in Fig. 11, where the theoretical spectrum generated from the implant depth and dose obtained by the ANN for sample 1 is superimposed onto the experimental data. While the overall agreement is good, some comments must be made. First, one should note that the spectrum was generated for nominal experimental conditions, in particular the nominal collected charge. Second, it is obvious that there is a large misfit below about channel 30. This is due to plural and multiple scattering, which is not taken into account and which leads to an increased yield at low energies [27]. Second, the slight misfit around channels 40–60 where the simulated spectrum is consistently above the data (4% on average) and where no plural and multiple scattering is expected, is due to the inaccuracy in charge collection. Also, the small peak observed in the experimental data around channel 37 is an oxygen signal due to the natural surface Si oxide.

Finally, and most importantly, the Ge peak is excellently reproduced, both in position and in size. The error in the implant depth committed by the ANN given in Table II,

65.4×10^{15} atom/cm 2 , is smaller than the depth resolution, which is in this case 153×10^{15} atom/cm 2 , as calculated with the DEPTH code [29]. Also, any inaccuracy in the energy calibration would lead to a slightly corrected depth value. As for the 2.39% error in the determination of the implanted dose, the statistical error corresponding to the Ge peak due to the limited collected charge and consequent limited Ge yield is 2.34% in this case (corresponding to 1818 counts in the Ge peak), to which one should add the 4% error in the nominal charge.

That is, the ANN is determining the implanted dose and depth as well as is possible given the constraints imposed by the physical model used in the calculation of theoretical RBS spectra, the accuracy in the experimental parameters, and the lack of *a priori* knowledge about the composition of the sample (in this case, the presence of a surface oxide). It should be noted that a careful analysis using NDF can overcome these constraints. A plural scattering correction can be introduced in NDF [39], the experimental parameters are automatically slightly adjusted by NDF to improve the internal consistency of the fit, and the surface oxide can also easily be included. All these corrections were taken into account in the values determined with NDF and given in Table II.

E. Influence of the RBS forward model

It is clear that the RBS forward model used has important consequences for the results obtained. To test the influence of different parameters in the results, we trained ANNs with the same architecture (1,100,80,50,0), but using training sets constructed using different assumptions. In all cases the errors in the training and test sets were comparable. We then applied the ANNs so built to the experimental data and calculated the mean absolute error for the implanted dose and depth. The results are given in Table III.

The average errors for the dose and depth obtained with the reference ANN, i.e., the one which includes all the effects including pileup correction and (statistical) Poisson noise and where the first 25 channels are ignored, are 2.56×10^{15} and 52.7×10^{15} atom/cm 2 , respectively. For an ANN trained with similar data, but where Poisson noise was not included (actually, exactly the same data except for the inclusion of the noise), the error in the dose improves signifi-

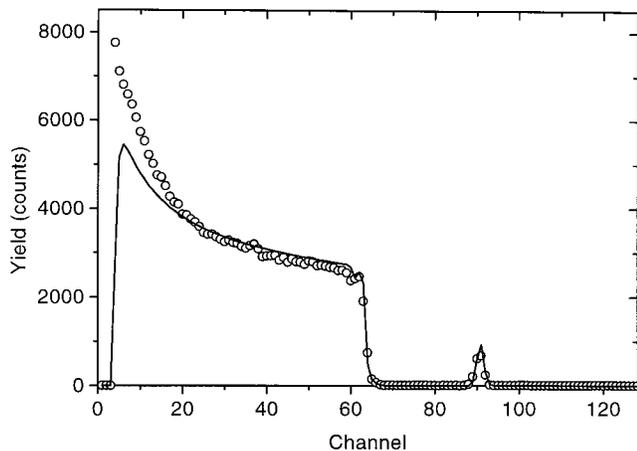


FIG. 11. Experimental data (open circles) and corresponding spectrum generated from the implant depth and dose obtained with the ANN for the nominal experimental conditions, for a given sample.

TABLE III. Average absolute error in the analysis of experimental data, obtained with the reference ANN and with the ANNs trained with generated data that did not include random Poisson noise, that did not include the pileup correction, and where only the first five channels were disregarded.

Absolute error (10^{15} atom/cm ²)	Reference ANN	No Poisson noise	No pileup correction	First five channels disregarded
Dose	2.36	1.22	3.45	5.65
Depth	52.7	64.1	68.3	106.2

cantly while the error in the implanted depth increases slightly. The improvement in the dose error can be understood by considering that ANNs recognize the shape and features of pictures, while the noise (experimental or calculated) is just a distortion of the original picture. By training the ANN with the undistorted signal, the ANN can more easily apprehend the important features in the data. From another point of view, one could say that also in traditional fitting methods the fitted curve is always purely theoretical with no noise introduced. As for the error in the implanted depth, it should be noted that the calculation of the depth depends strongly on the position of the signal and only weakly on its exact shape, and hence the depth error changes only slightly.

We also trained an ANN using data with Poisson noise and ignoring the first 25 channels, but where the pileup correction was not applied. Again, the depth error does not change very much, while the error in the determination of the dose is significantly worse. The reason is that the main effect of the pileup correction is to reduce or increase the yield in different parts of the spectrum: where there were originally no (or little) counts, pileup will lead to an increase in the yield, and where there were originally many counts, the yield will decrease. So around the Ge peak the yield increases due to pileup while in the peak itself it decreases, leading to an extra source of error when pileup is not considered.

Finally, we trained an ANN using exactly the same data as in the reference ANN, but considering the whole spectra (that is, disregarding the first 5 instead of the first 25 channels as in the reference ANN) both in the training process and in the analysis of the experimental data. The error in both the dose and depth increases by a factor of about 2. This is due to the strong distortion in the experimental data that is introduced by plural scattering. On the one hand, the Ge dose is given by the yield in the Ge peak, but the absolute normalization is given by the height of the Si signal. The increase in the Si yield due to plural scattering is hence bound to lead to an increase in the dose error. On the other hand, the determination of the depth is based on the position of the Ge signal relative to the position where surface Ge would be. The difference is determined by the stopping power of the incident He beam in Si, which also strongly influences the shape of the Si signal, particularly at low energies where the stopping power is a strong function of the beam energy. Hence one of the features of the data that the ANN can use to determine the Ge depth is the shape of the Si signal, and the plural scattering distortion will lead to an increased error.

V. CONCLUSIONS

As far as we are aware, the artificial neural network algorithm demonstrated here is the first method that can analyze

RBS data instantaneously. Other methods developed that are able to perform automated analysis of RBS data, such as Bayesian inference or simulated annealing, require some time to perform the analysis, as well as some degree of understanding of the physics behind RBS and of the mathematics behind the algorithms. In contrast, the ANN algorithm presented here is essentially a push-button black box, which opens the doors to the integration of RBS and other ion beam analysis techniques in the production line.

The algorithm developed can be applied to a single system, namely, implants of Ge in Si, albeit in a very wide range of implanted dose and depth, as well as in a very wide range of experimental conditions. For each different system to be studied, a new ANN must be built and trained. While this is clearly a shortcoming, many applications require the analysis of a large amount of similar samples. This is the case, for instance, in quality control or in batch production of given systems.

The reliability of the ANN developed was investigated. It was shown that there exist two well-defined classes of spectra that the ANN is not able to analyze correctly. These two classes, however, correspond to cases that are difficult to analyze even using traditional methods. They correspond to either very large implant depths, where the Ge and Si signals overlap and are difficult to separate, or to very small doses, where it is difficult to distinguish the Ge signal from the background. The experimentalist's solution to solving the problem would be, in one case, to change the angle of incidence or to increase the beam energy and, in the latter case, to increase the beam fluence. While this normally requires the presence of a highly trained experimentalist during the measurement, implementation of feedback from the ANN to the experimental setup would lead to on-line automated optimization of the experimental conditions for each given sample. Further research in this direction is being developed.

The ANN developed was applied to real experimental data with excellent results. The error in the implanted dose was within the experimental error given by the limited counting statistics, and the error in the implanted depth was within the depth resolution of the technique. This means that the results given by the ANN can be taken directly as such or used as the initial guess in a fast local optimization algorithm.

ANNs are often trained using real data. For RBS, that would require measuring thousands or tens of thousand of spectra and analyzing them by traditional methods. This is, in practice, not feasible. The alternative found was to train the ANN with theoretically generated spectra. On the one hand, this was validated by the successful application of the ANN so trained to real experimental data. We nevertheless

studied the influence of the theoretical RBS model used to generate the spectra on the ANN performance. The conclusion is that to obtain maximum accuracy of the results all relevant effects should be included in the model. While pileup can be easily taken into account, the effect of plural and multiple scattering cannot be calculated efficiently and accurately enough for the purposes of this study. This means that the low-energy region of the data, which can be strongly

affected by plural and multiple scattering, cannot be used in the analysis of experimental data with the ANN developed here. This does not affect the particular system studied, since the Ge signal is located in the high-energy region of the data. However, future developments will have to include the effect of multiple and plural scattering, either by utilization of some experimental data in the training or by devising an efficient way of calculating it.

-
- [1] *Handbook of Modern Ion Beam Materials Analysis*, edited by J. R. Tesmer and M. Nastasi (Materials Research Society, MRS, Pittsburgh, 1995).
- [2] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).
- [3] Emile Aarts and Jan Korst, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing* (Wiley, Chichester, 1989).
- [4] N. P. Barradas, C. Jeynes, and R. P. Webb, *Appl. Phys. Lett.* **71**, 291 (1997).
- [5] N. P. Barradas, P. K. Marriott, C. Jeynes, and R. P. Webb, *Nucl. Instrum. Methods Phys. Res. B* **136–138**, 1157 (1998).
- [6] The Data Furnace: Manual for version v1.0 (October 1997), update to version v2.1 (March 1998), and introduction to version v6 (October 1999), University of Surrey, <http://www.ee.surrey.ac.uk/Research/SCRIBA/ndf/>.
- [7] N. P. Barradas, C. Jeynes, R. P. Webb, U. Kreissig, and R. Gröttschel, *Nucl. Instrum. Methods Phys. Res. B* **149**, 233 (1999).
- [8] N. P. Barradas and R. Smith, *J. Phys. D* **32**, 2964 (1999).
- [9] Christopher M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, Oxford, 1995).
- [10] L. Duponchel, C. Ruckebusch, J. P. Huvenne, and P. Legrand, *J. Near Infrared Spectrosc.* **7**, 155 (1999).
- [11] N. Aras, B. J. Oommen, and I. K. Altinel, *Neural Networks* **12**, 1273 (1999).
- [12] T. Smith, S. Carr-Lopez, and S. Phillips, *J. Am. Geriatr. Soc.* **47**, 174 (1999).
- [13] V. M. Prozesky, J. Padayachee, R. Fischer, W. von der Linden, V. Dose, and R. A. Weller, *Nucl. Instrum. Methods Phys. Res. B* **136–138**, 1146 (1998).
- [14] N. P. Barradas, J. L. Keddie, and R. Sackin, *Phys. Rev. E* **59**, 6138 (1999).
- [15] A. D. Dane, H. A. van Sprang, and L. M. C. Buydens, *Anal. Chem.* **71**, 4580 (1999).
- [16] A. Vieira and N. P. Barradas, *Nucl. Instrum. Methods Phys. Res. B* **170**, 235 (2000).
- [17] R. P. G. Karunasiri, J. S. Park, and K. L. Lang, *Appl. Phys. Lett.* **59**, 2588 (1991).
- [18] H. Geiger, *Proc. R. Soc. London, Ser. A* **83**, 492 (1909).
- [19] H. Geiger, E. Marsden, *Proc. R. Soc. London, Ser. A* **82**, 495 (1909).
- [20] E. Rutherford, *Philos. Mag.* **21**, 669 (1911).
- [21] H. Geiger, E. Marsden, *Philos. Mag.* **25**, 604 (1913).
- [22] J. L'Ecuyer, J. A. Davies, and N. Matsunami, *Nucl. Instrum. Methods* **160**, 337 (1979).
- [23] J. F. Ziegler, J. P. Biersack, and U. Littmark, *Stopping and Ranges of Ions in Solids* (Pergamon, New York, 1985).
- [24] C. Jeynes, N. P. Barradas, P. K. Marriott, M. Jenkin, E. Wendler, G. Boudreault, and R. P. Webb (unpublished).
- [25] E. Kótai, *Nucl. Instrum. Methods Phys. Res. B* **85**, 588 (1994).
- [26] M. M. Li and D. J. O'Connor, *Nucl. Instrum. Methods Phys. Res. B* **149**, 460 (1999).
- [27] A. Weber and H. Mommsen, *Nucl. Instrum. Methods Phys. Res.* **204**, 559 (1983).
- [28] P. Bauer, E. Steinbauer, and J. P. Biersack, *Nucl. Instrum. Methods Phys. Res. B* **64**, 711 (1992).
- [29] E. Szilágyi, F. Pászti, and G. Amsel, *Nucl. Instrum. Methods Phys. Res. B* **100**, 103 (1995).
- [30] E. Szilágyi, *Nucl. Instrum. Methods Phys. Res. B* **161–163**, 37 (2000).
- [31] B. Cheng and D. M. Titterington, *Stat. Sci.* **9**, 2 (1994).
- [32] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, edited by D. E. Rumelhart and J. L. McClelland (MIT Press, Cambridge, MA, 1986), Vol. 1, p. 318.
- [33] P. Baldi and K. Hornik, *Neural Networks* **2**, 53 (1989).
- [34] G. E. Hinton, in *Proceedings of the PARLE Conference on Parallel Architectures and Languages Europe*, edited by J. W. De Bakker, A. J. Nijman, and P. C. Treleaven (Springer-Verlag, Berlin, 1987), p. 1.
- [35] Melanie Mitchell, *An Introduction to Genetic Algorithms* (MIT Press, Cambridge, MA, 1998), pp. 65–79.
- [36] C. Jeynes, Z. H. Jafri, R. P. Webb, A. C. Kimber, and M. J. Ashwin, *Surf. Interface Anal.* **25**, 254 (1997).
- [37] A. F. Gurbich, *Nucl. Instrum. Methods Phys. Res. B* **364**, 496 (1995).
- [38] N. P. Barradas, A. P. Knights, C. Jeynes, O. A. Mironov, T. J. Grasby, and E. H. C. Parker, *Phys. Rev. B* **59**, 5097 (1999).
- [39] N. P. Barradas, C. Jeynes, and S. M. Jackson, *Nucl. Instrum. Methods Phys. Res. B* **136–138**, 1168 (1998).