Downloaded from rsta.royalsocietypublishing.org

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Noise, Resolution and Entropy in Sputter Profiling

M. G. Dowsett and R. Collins

Phil. Trans. R. Soc. Lond. A 1996 **354**, 2713-2729
doi: 10.1098/rsta.1996.0125

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

To subscribe to *Phil. Trans. R. Soc. Lond. A* go to:
http://rsta.royalsocietypublishing.org/subscriptions

This journal is © 1996 The Royal Society

Downloaded from rsta.royalsocietypublishing.org

Noise, resolution and entropy in sputter profiling

By M. G. Dowsett[1] and R. Collins[2]

[1]*Advanced SIMS Projects Group, Department of Physics, University of Warwick, Coventry CV4 7AL, UK*

[2]*Department of Physics, University of York, Heslington, York YO1 5DD, UK*

Applications of semiconductor sputter profiling in the near future will demand nm or even sub-nm depth resolution, i.e. the ability to distinguish between adjacent features in a concentration profile which are separated by 1 nm or less (not just the ability to achieve a 16–84% width of 1 nm on some edge transient). Recent developments in SIMS indicate that sputter profiling can meet this challenge—provided that instruments with high current microfocus beams less than 1 keV in energy become widely available, and data processing methods are developed to cope with the results. In this paper we discuss the nature of SIMS sputter profile data in detail and explore the consequences of noise, finite sample interval and the destruction of information by processes such as atomic mixing. The use of empirically determined response functions is discussed. We use simple arguments to show that convolution is only a valid model for atomic mixing in the low concentration limit, and contrast the use of Fourier transform and maximum entropy methods for deconvolution in these circumstances. A depth resolution of 1 nm is demonstrated without deconvolution, showing that the technique has real potential for depth resolution on the scale of 1 atomic plane.

1. Introduction

Secondary ion mass spectrometry (SIMS) is a powerful technique for analysing small volumes of almost any non-volatile material. A monoenergetic primary ion beam, typically in the energy range 0.2–15 keV, bombards the sample *in vacuo* and sputtered secondary ions, characteristic of the sample, are analysed by mass spectrometry. Information at high spatial resolution can be obtained by controlling the shape of the analyte volume. For example, scanning the primary ion beam in a raster of 1–100s μm^2 and collecting data for each pixel in the scan, provides a surface chemical image with a lateral resolution which can be better than 50 nm (Levi-Seti *et al.* 1991). Repeating this process sequentially on the same area, produces a flat bottomed crater (ideally) revealing successively deeper material. A high depth resolution chemical depth profile can be extracted from data collected in this way (for example, Dowsett *et al.* 1994*a*). (Recently, through the use of primary beam energies below 1 keV, depth resolutions *ca.* 1 nm have been achieved as we shall demonstrate later.) Data may be stored as a 'stack' of images so that profiles and sections may be extracted *a posteriori* (Rüdenauer 1994). Alternatively, if the primary dose is kept below 10^{13} ions cm^{-2} during the complete analysis, SIMS has high surface specificity

Phil. Trans. R. Soc. Lond. A (1996) **354**, 2713–2729
Printed in Great Britain

© 1996 The Royal Society
TEX Paper

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions: Mathematical, Physical and Engineering Sciences.* www.jstor.org

Downloaded from rsta.royalsocietypublishing.org

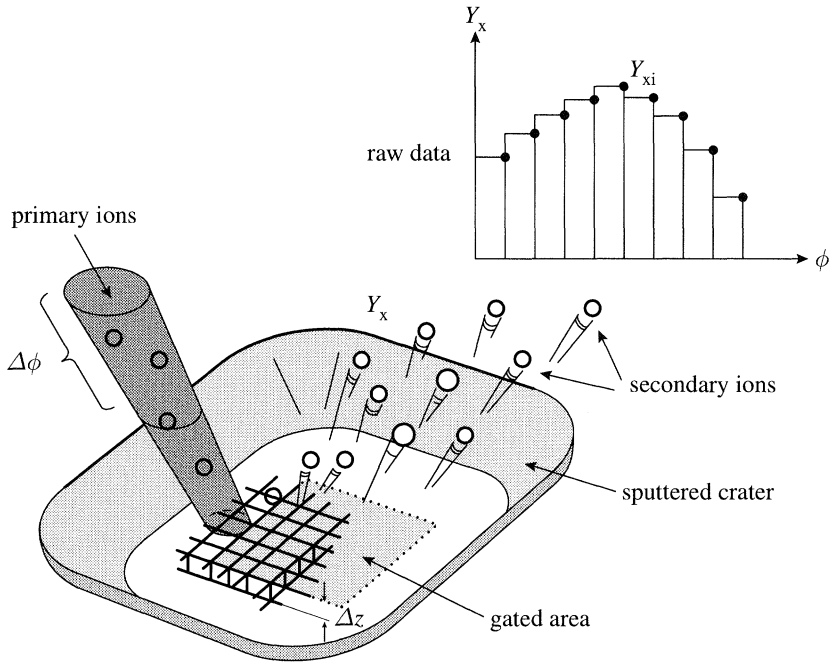

Figure 1. The origin of the data in a SIMS depth profile. The dots on the histogram show where data points would be plotted in practice.

(static SIMS) and can be used to obtain mass spectra representative of the chemistry of the top monolayer or so (Briggs 1992).

This paper discusses the quantification of data obtained using SIMS as a depth profiling technique at the limits of depth resolution. However, most of the arguments will apply to all SIMS experiments, and also to other analytical techniques. We begin with a description of the SIMS data collection process, and the data themselves.

2. The data

The basic datum Y acquired in a SIMS measurement consists of a count of the number of secondary ions arriving at a detector for an incremental dose $\Delta\phi$ of primary ions striking the sample. The type of analysis obtained out of the examples described above depends on which instrumental parameters are varied between the acquisition of one datum and the next. For example, in the simplest imaginable depth profile for one species (say an impurity element X in a matrix A), the primary beam will be scanned over a small (typically less than 0.2 mm^2) area on the sample in a raster pattern of 256×256 overlapping pixels while the mass spectrometer is tuned to species X (figure 1). While the beam resides in the central 10% or so of the area (the gate), a count Y_X is accumulated. This cycle of events is repeated many times to obtain a set of N ordinates for X. As shown in figure 1, a histogram is acquired for Y_X versus ϕ, and if the current of X secondary ions contains detailed structure within the dose increment $\Delta\phi$ we have no way of knowing. If we add a second mass channel for the matrix A, and alternate between sampling Y_X and Y_A, keeping the sampling dose at $\Delta\phi$, any detail on a scale less than $2\Delta\phi$ is now lost to us, and so on. Provided the primary beam current and current density are stable, the dose

Downloaded from rsta.royalsocietypublishing.org

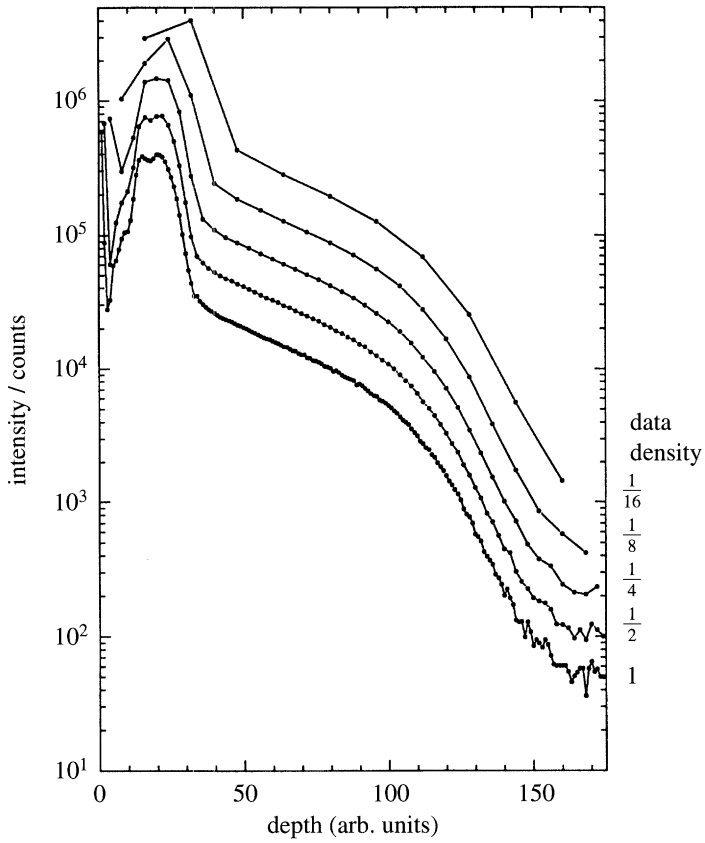

Figure 2. SIMS depth profile of a 30 keV 10^{16} cm^{-2} annealed boron implant into silicon. The five increasingly coarse sampling intervals illustrate the effects of undersampling in the raw data.

increment may be replaced by the time increment taken for one complete raster, Δt. With this proviso in mind, ϕ and t may be taken as interchangeable throughout the rest of this paper. For the ideal measurement in a single channel profile, quantification of the ith ordinate for X, Y_{X_i} would consist of converting the dose value of $i\Delta\phi$ to a depth and the Y_{X_i} to a concentration.

The SIMS data have the following features in common with all experimental data. Firstly, they comprise a finite set of measurements (separated by gaps of $\Delta\phi$ and accumulated over doses of $\Delta\phi$)—they are underdefined. Nevertheless, our inclination is to treat the data as if it were continuous, and we forget that our knowledge of what goes on in the gaps is based on inference—educated guesswork about how we expect the system to behave (e.g. smooth, continuous, etc.). Figure 2 shows SIMS profiles of the same data at five sampling intervals to illustrate the sort of errors that can be generated by underdefinition. The data are from a heat treated boron implant into silicon. The boron had (unexpectedly) precipitated in high concentration regions giving rise to the structure visible in the high sampling density profile. When the data were originally collected at low sampling rates, the irregular appearance of the profile led to the incorrect inference that either the instrument was malfunctioning, or the sample was charging during analysis.

Secondly, the data contain fluctuations—if we repeat the profile under the 'same' experimental conditions, the ith ordinate is very likely to have a different value each

Downloaded from rsta.royalsocietypublishing.org

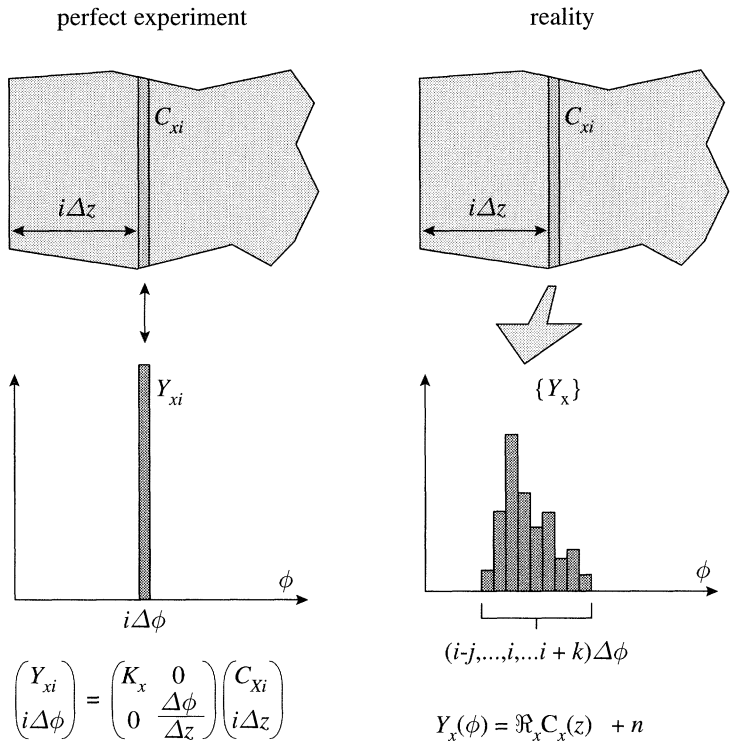

Figure 3. The difference between a perfect (invertable) measurement process, and a real SIMS depth profile measurement involving permanent information loss because of ion beam induced broadening effects.

time. Thirdly, the experimental process will have contributed something to the measurement which results in (permanent) loss of information—in this paper we loosely refer to this as blurring (we do not wish to imply that the process is necessarily linear). In sputter profiling this arises because the deposition of both energy and atoms from the primary beam in the near-surface region of the sample causes internal migration of the original chemical distribution before it is sputtered (ion beam induced mass transport). Amongst the processes involved are three which are inevitable—micro pitting (due to individual ion impacts) (Wang *et al.* 1995), random cascade mixing (Littmaack & Hofer 1980) and Kirkendall effects (Collins *et al.* 1988)—and several which occur for certain combinations of experimental conditions, e.g. chemical segregation (Boudewijn *et al.* 1984). In SIMS there is an additional effect which may be dominant: the probability of ejecting a secondary ion as a result of a primary ion impact is nonlinearly dependent on the local surface chemistry (the SIMS matrix effect) (Wittmaack 1980). Here, we are concerned with the loss of information due to a combination of these processes, and do not distinguish between them. If in a perfect experiment we could ascribe all the count in the ith ordinate Y_i at a dose $i\Delta\phi$ to a chemical concentration C_i in a depth increment Δz at a depth $i\Delta z$, then in our real experiment, Y_i contains contributions from $C_{i-j} \ldots C_i \ldots C_{i+k}$, where the range j to k depends on the nature of the blurring (figure 3).

Finally, the data will contain statistical noise from particle counting and interference from the cosmic ray background, even if the SIMS instrument were otherwise perfect.

Downloaded from rsta.royalsocietypublishing.org

There are five extra things to note about the data. (i) Neither the noise, nor the interference will have been subjected to the blurring process. They can therefore contribute sharper local changes in Y than would be possible otherwise; they add high frequencies to the Fourier transform of the data. (ii) There is currently no useful theoretical model which can accurately predict the effects of the blurring process on a given chemical distribution (although such models are under development). If it is required, the blurring must be determined empirically, for, say, atomically thin layers of B in A (so-called delta layers which can be produced by various deposition methods such as molecular beam epitaxy (MBE)). The response function obtained in this way is itself underdefined and noisy and hence contains incomplete information on the blurring effect. This and other limitations will be discussed later. (iii) SIMS depth profile data have a dynamic range which usually exceeds four orders of magnitude, and can exceed six. (iv) It is everywhere true that $C_i \geqslant 0$ and $Y_i \geqslant 0$; negative concentrations and negative counts are physically unrealistic. (v) From N measurements Y_1 to Y_N we normally require N estimates C_1 to C_N of the true concentration profile. In any quantification scheme, we can therefore construct as many equations as there are unknowns.

These are the data we must quantify. SIMS depth profiling is, perhaps, the simplest variant of the SIMS technique, and gives ample warning to the analyst of when the measurements are unquantifiable (i.e. the set of counts $\{Y\}$ cannot be converted accurately to a set of concentrations $\{C\}$ by any simple means). Understanding this one-dimensional case therefore forms a good foundation on which to build necessarily more sophisticated quantification schemes for two- and three-dimensional data. Although the same difficulties are present in SIMS imaging, for example, there always seems to be a desire to interpret contrast directly in terms of concentration. Unfortunately, it is quite possible for an increase in image intensity of a factor of 100 to correspond to a drop in concentration of the analyte of a factor of 10—all that is required is the presence locally of an ion yield enhancing species such as oxygen. Although analysts may be aware of this problem, in the author's experience, the end users of the data are often not. Such effects are more obvious where they occur in one-dimensional information.

One conclusion that might be reached on the basis of the above description of the data is that some deconvolution method could form part of the data processing. There are two basic deconvolution methods: inverse methods based on Fourier transform properties of the convolution integral, and forward methods based on trial and error which are also applicable to more general inversion problems. These are outlined later on. For now, we state that: inverse Fourier transform (FT) methods (applied to data sets as described above) *always* produce the wrong answer. Forward methods, such as a method which we have developed based on maximum entropy (MaxEnt) (Allen *et al.* 1993) *might* have produced the right answer—*but how could one tell?* Both methods are tools which may under some circumstances provide better *estimates* of the true concentration profile than can be obtained from simple processing of the raw data. In order to apply either of them it is necessary to have a model of the way the blurring transforms the true distribution. For the FT method, the model is explicitly convolution—*and if convolution is an inappropriate model, e.g. the process is nonlinear as described in § 3 below, then FT is an inappropriate method.* For the MaxEnt case, any forward model may be adopted, although for a nonlinear process neither the MaxEnt (nor any other) solution may be unique.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Downloaded from rsta.royalsocietypublishing.org

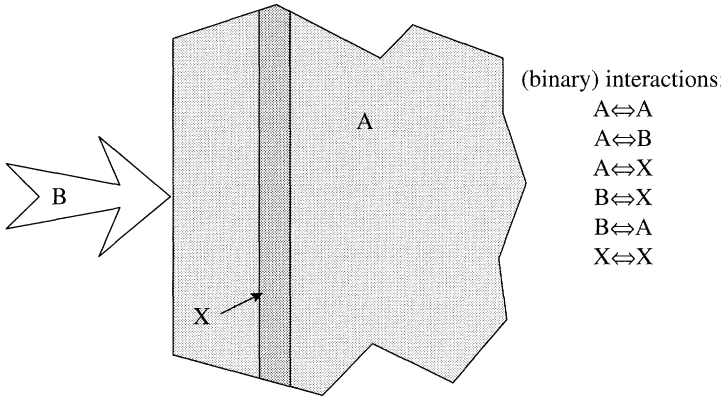

Figure 4. Possible binary interactions in a SIMS depth profile for species X in matrix A using primary ions B. If X is sufficiently dilute to render X⇔X interactions unlikely, and to have no effect on the interaction of B with A, then the system should be linear.

3. Linearity or nonlinearity in the process model

The models which exist for ion beam induced mass transport (in the high fluence regime) involve analytic (e.g. Littmaack & Hofer 1980) or numerical (Collins 1986) treatments of the mass currents for each species involved. The models are nonlinear. There are no generally useful models for the ion emission process at present, but measurements show for concentrations exceeding 1% the ion yield is not necessarily proportional to, nor even monotonic with, concentration. Returning to the simple experiment in the introduction, we seek to measure the impurity distribution for X in a matrix of A using primary ions of type B (figure 4).

The yield of X ions will depend upon exactly how B interacts with A and X in the presence of X or A. It seems reasonable that, if the concentration of X is sufficiently low, the ion beam induced mass transport and ionization probability effects (for X) will be dominated by interactions of the type A⇔X and B⇔X—i.e. X⇔X interactions are too improbable to play a significant role. Similarly, there will be some concentration below which X does not significantly affect the way B combines with the sample surface. Under these circumstances, changes in Y_{X} will be proportionally related to changes in its concentration C_{X}. This critical concentration $C_{\mathrm{dil\ X}}$ is called the dilute limit. Above the dilute limit, or where X is sought in a multi-matrix sample $(\mathrm{A}_1 \ldots \mathrm{A}_n)$, where each matrix volume (thickness in a depth profile) is too small for the A⇔B interactions to come into equilibrium, the SIMS depth profiling process is described by

$$Y_{\mathrm{X}}(\phi) = \mathcal{R}_{\mathrm{X}} C_{\mathrm{X}}(z) + n, \tag{3.1}$$

where $\mathcal{R}_{\mathrm{X}}$ is a nonlinear operator describing all the physical effects, $C_{\mathrm{X}}(z)$ is the true concentration profile of X as a function of depth and n is additive noise (statistical noise plus interference). Additive is here taken to imply that the noise is effectively added *after* the blurring process as described in the introduction. Such phenomena as fluctuation in the primary beam current, statistical fluctuations and interference in the counting system may all be accumulated into n.

At the other extreme, below the dilute limit and in the quantification of depth profiles for applications requiring low depth resolution, both the blurring and the noise are ignored and the process model is (implicitly) assumed to be an invertable

Downloaded from rsta.royalsocietypublishing.org

mapping:

$$Y_{\mathrm{X}}(\phi) = \begin{pmatrix} Y_{\mathrm{X}_i} \\ i\Delta\phi \end{pmatrix} = MC_{\mathrm{X}}(z) = \begin{pmatrix} K_{\mathrm{X}} & 0 \\ 0 & \Delta\phi/\Delta z \end{pmatrix} \begin{pmatrix} C_{\mathrm{X}_i} \\ i\Delta z \end{pmatrix}, \tag{3.2}$$

where $K_{\mathrm{X}} = \tau_{\mathrm{X}} A \Delta z$ (τ_{X} is the useful yield for X, A is the analysed area) and $\Delta\phi/\Delta z$ is the erosion rate. If the measurement involves more than one mass channel, then $\Delta\phi$ needs to be modified appropriately. This equation is evidently the model for a *perfect* SIMS experiment, so we can write

$$Y_{\mathrm{X\ perf}}(\phi) = MC_{\mathrm{X}}(z). \tag{3.3}$$

Then, a more realistic linear model below the dilute limit would be

$$Y_{\mathrm{X}}(\phi) = n + \int \mathrm{d}\phi'\, R_{\mathrm{X}}(\phi, \phi') Y_{\mathrm{X\ perf}}(\phi'), \tag{3.4}$$

where $R_{\mathrm{X}}(\phi, \phi')$ is a superposable response function which varies, in general, with ϕ (or z). If it can be established *by experiment* that R_{X} is invariant with ϕ (or z) then equation (3.4) reduces to a convolution integral:

$$Y_{\mathrm{X}}(\phi) = n + \int \mathrm{d}\phi'\, R_{\mathrm{X}}(\phi - \phi') Y_{\mathrm{X\ perf}}(\phi'). \tag{3.5}$$

Where convolution is an acceptable model for the SIMS depth profile process, deconvolution is the correct method of quantification. One must then decide how deconvolution is to be done, and what limitations are imposed by the finite sample and the presence of n.

Unfortunately, it is commonly *assumed* that convolution is an acceptable model for sputter profiling in general, and SIMS in particular. The fact that deconvolution algorithms produce plausible results which appear 'better' in some sense than the original data leads people to apply them uncritically. It is important to understand that a hand drawn 'better' estimate of what the data represent may have more validity than anything gained through the spurious application of a mathematical process. It is essential to establish necessary and sufficient conditions for the validity of the convolution model *for the matrix–impurity–beam* combination of interest. This is not easy to do, but the conditions may be summarized as follows: it is necessary that the shape of the response function R is independent of both the depth and concentration of the feature on which it is measured; it is sufficient that the shape of Y_{meas} for a thick uniform layer (the superposition of a set of R) is everywhere independent of its concentration C_{max}, and that

$$\int_{\phi} Y_{\mathrm{meas}}(\phi)\, \mathrm{d}\phi \propto C_{\mathrm{max}}. \tag{3.6}$$

These conditions are sufficient because they exclude the possibility of the response function shape remaining constant beyond the dilute limit because of cascade dilution (Williams & Baker 1980).

4. Forward and inverse methods for the estimation of $C(z)$

Forward methods are robust and aim to simulate a noise free measurement process and compare the results with the measured data. One takes a guess $C_{\mathrm{trial}}(z)$ (often

Downloaded from rsta.royalsocietypublishing.org

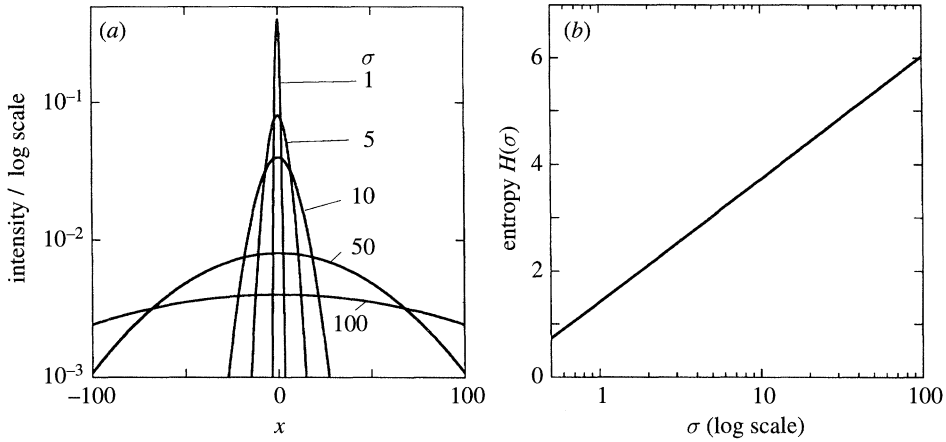

Figure 5. (*a*) The normalized Gaussian distribution $f(x,\sigma) = (1/\sqrt{2\pi}\sigma)\mathrm{e}^{-x^2/2\sigma^2}$ plotted with σ as a parameter and (*b*) its entropy $H(\sigma) = 0.5 + \ln(\sqrt{2\pi}\sigma)$. As the data become more featureless (broader Gaussian) the entropy increases.

featureless) at the true chemical profile and operates on it with the model (equations (3.1), (3.4) or (3.5) with $n = 0$) to produce a synthetic data set $Y_{\mathrm{trial}}(\phi)$. One constructs a potential function V consisting of a linear combination of a measure A of how closely the $Y_{\mathrm{trial}}(\phi)$ agrees with the measured data Y_{meas}, and an independent measure B of some desirable characteristic of the C_{trial} (for example, Press *et al.* 1992*a*). One then seeks the best estimate of the truth C_{est} by modifying $C_{\mathrm{trial}}(z)$ to maximize V, where the gradient of V can be used to control the search. The misfit statistic χ^2 may be used for A:

$$\chi^2 = \sum_N \frac{[Y_{\mathrm{meas}\ i} - Y_{\mathrm{trial}\ i}]^2}{\sigma_i^2}, \tag{4.1}$$

where σ_i is the expectation value of the noise on $Y_{\mathrm{meas}\ i}$. The range of choice in C_{trial} is constrained by B which is known as a regularization parameter. One of its tasks is to prevent transfer of the noise on Y_{meas} back to C_{trial}. In SIMS it is important that the method selected does not directly constrain the local gradient (the sharpness) of C_{trial}, since this would be to arbitrarily reject possible solutions to the inversion problem (but see also § 6). At the same time, structure should not appear in C_{trial} for which there is insufficient evidence in Y_{meas}. A possible choice for A is therefore the information or Shannon entropy H where (for example, Gull 1994)

$$H = -\sum_N C_i \log C_i \quad \text{and} \quad \sum_N C_i = \text{constant (usually 1)}, \tag{4.2}$$

and the C_i are the ordinates of C_{trial}. (This definition of the entropy applies to measurements of high statistical significance where, in effect, $Y_{\mathrm{meas}} \gg 1$. It may not be the most appropriate form for recovering data over several orders of magnitude or from very noisy data.) This is the basis of the MaxEnt method (Skilling & Gull 1985). Figure 5 gives a simple demonstration of the fact that increasing the entropy drives a distribution towards featurelessness.

The potential function becomes:

$$V = \alpha H - \chi^2. \tag{4.3}$$

The Lagrange multiplier α controls the balance between entropy and misfit. Finding

Downloaded from rsta.royalsocietypublishing.org

α is part of the optimization procedure. It may be done by 'chopping' (Skilling & Gull 1985) or trial and error subject to $\chi^2 = N$ (Allen 1994). These methods require maximization of V for each value of α tested— rather slow because of the nested loop structure required. An alternative and faster method involves a self-optimizing form of α which converges as V is maximized (Myrheim & Rue 1992). In either case, the overall process may be described as *'find the most featureless* C_{trial}, C_{est} *which when convolved with* R *gives a simulated result consistent with the measured data'*, i.e.

$$R * MC_{\text{est}} = Y_{\text{meas}}, \quad \text{subject to } \chi^2 \leqslant N. \tag{4.4}$$

The form of the entropy automatically ensures positivity in the recovered information, since $\partial V/\partial H \to \infty$ as $C \to 0$. A powerful argument often put forward for the use of MaxEnt is its 'objectivity' or freedom from bias. Whether or not this process can be considered objective in practice depends very much on having a correct noise model and a 'correct' value for α.

Inverse methods are based on the Fourier transform properties of the convolution integral:

$$\tilde{y}(k) = \tilde{r}(k) \times \tilde{y}_{\text{perf}}(k), \tag{4.5}$$

where $\tilde{y}(k)$ is the Fourier transform of $Y(\phi)$, etc. Hence one might write

$$C_{\text{est}}(z) = M^{-1} \int \mathrm{d}k \frac{\tilde{y}(k)}{\tilde{r}(k)} \mathrm{e}^{-\mathrm{i}kz}. \tag{4.6}$$

Unfortunately, equations (4.5) and (4.6) leave out the effects of underdefinition and noise. In practice, $\tilde{y}(k)$ is the transform of $Y(\phi) + n$. The noise has not been blurred by transformation through R and will contain finite amplitude for k beyond the range of $\tilde{r}$. Using equation (4.6) unmodified produces extreme noise amplification and physically unrealistic negative regions. The effect of trying to deblurr data which contains features too narrow to have been produced by the blurring process may be demonstrated analytically. We take a small 'noise' element

$$n(\phi) = \frac{1}{\sigma_n \sqrt{2\pi}} \mathrm{e}^{-\phi^2/2\sigma_n^2}, \tag{4.7}$$

represented by the Gaussian distribution, and a response

$$R(\phi) = \frac{1}{\sigma_R \sqrt{2\pi}} \mathrm{e}^{-\phi^2/2\sigma_R^2}, \tag{4.8}$$

where their respective Fourier transforms are $\tilde{n}(k) = \mathrm{e}^{-k^2\sigma_n^2/2}$ and $\tilde{r}(k) = \mathrm{e}^{-k^2\sigma_R^2/2}$. Recovering the local chemistry according to equation (4.6), we have that

$$\tilde{c}(k) = \mathrm{e}^{-k^2(\sigma_n^2 - \sigma_R^2)/2}, \tag{4.9}$$

and so

$$C(z) = M^{-1} \begin{cases} \dfrac{1}{\sqrt{2\pi(\sigma_n^2 - \sigma_R^2)}} \exp[-z^2/2(\sigma_n^2 - \sigma_R^2)] & \text{for } \sigma_n > \sigma_R, \\ \delta(z), \quad \text{for } \sigma_n = \sigma_R, \\ \text{no real solution otherwise.} \end{cases} \tag{4.10}$$

One therefore needs to smooth or filter Y_{meas} until it contains no features narrower

Downloaded from rsta.royalsocietypublishing.org

than the response (i.e. its normalized frequency envelope lies everywhere within the envelope of the response). The recovered data are then

$$C_{\mathrm{est}} = M^{-1} \int \mathrm{d}k \frac{\tilde{y}(k) \times \tilde{s}(k)}{\tilde{r}(k)} \mathrm{e}^{-\mathrm{i}kz}, \tag{4.11}$$

where the $\tilde{s}(k)$ represents the smoothing. Since this $C_{\mathrm{est}}(z)$ can never satisfy equation (4.4), the forward method is better in at least this respect. Although forward methods tend to take more computing time than inverse methods, (minutes compared with less than 100 ms for MaxEnt and fast Fourier transform (FFT) based algorithms (e.g. Press *et al.* 1992*b*) respectively on a Pentium P90 PC) the time is significantly shorter than that taken in the data acquisition.

5. Limitations imposed by underdefinition

Given N points for the distribution $\{Y_{\mathrm{X}}\}$ we can only recover at best N independent points for the estimate of concentration $\{C_{\mathrm{X}}\}$. The immediate question which arises is how large does N need to be, or how small must Δz (and thus $\Delta\phi$) be. Note that the requirement to achieve a certain statistical precision in a measurement will mean that $\Delta\phi$ cannot be made arbitrarily small, i.e. a volume of material sufficient to give a significant count must be consumed to get each Y ordinate. The narrowest $C_{\mathrm{X}}(z)$ which could exist would be a distribution of X on a single atomic plane $\{0, \ldots, 0, C_{xi}, 0, \ldots, 0\}$ (i.e. a delta layer)—say 0.3 nm wide. (This is obviously a slightly idealized description—real samples would have, at best, a very narrow Gaussian distribution about a surface containing atomic steps.) Any broader distribution can be regarded as a stack of deltas—so here is the ultimate resolution limit. Suppose we have a sample which consists of a superlattice of repeated deltas $\{C_{x1}, 0, \ldots, C_{xi}, 0, \ldots, C_{xN}, 0\}$. From the Nyquist condition, we could *just* reconstruct a sinusoidal $C_{\mathrm{X}}(z)$ with this half period if Δz =0.3 nm (through inversion of equations (3.1), (3.4) or (3.5) with $n = 0$). To see what is required to reconstruct the profile more fully we need only look at the Fourier series for the superlattice (i.e. the Fourier series for a square wave offset by addition of the amplitude C_0 so that $C_{\mathrm{X}}(z)$ varies between zero and $2C_0$):

$$C_{\mathrm{X}}(z) = C_0 \left(1 + \sum_n \frac{1}{2n-1} \sin(2n-1) \frac{z\pi}{0.6} \right), \tag{5.1}$$

where z is depth in nm. Figure 6 shows the best profiles which could be reconstructed using FT methods (in the absence of noise) if $\Delta\phi$ corresponded to depth intervals of 0.3, 0.1, and 0.06 nm. The recovered data would sample the curves shown at the same pitch, but at a phase which depended on the depth of the features below the surface.

Forward methods also suffer from undersampling of the data. Figure 7 shows a MaxEnt reconstruction of a profile of a set of three boron delta layers in silicon taken with a high energy (significantly blurring) primary beam, compared with a relatively low energy profile of the same sample. The two profiles do not superimpose. A high energy probe usually contains more current and profiles faster than a low energy one, unless the experimenter is prepared to control the erosion rate (often at the expense of signal). For each individual layer in the high energy profile, the search criteria are satisfied within *ca.* $\pm\Delta\phi$ of the more accurate position measured by the low energy

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Downloaded from rsta.royalsocietypublishing.org

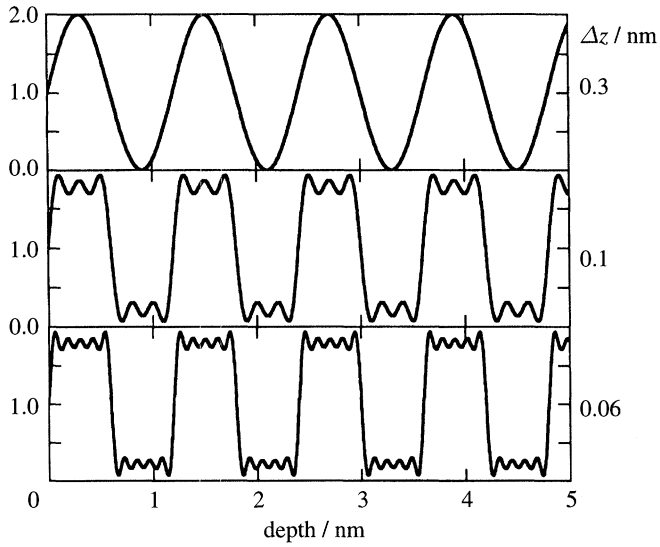

Figure 6. The best reconstructions of the ultimate multilayer sample described in the text using frequencies up to the Nyquist limit for sampling intervals of 0.3, 0.1, and 0.06 nm.

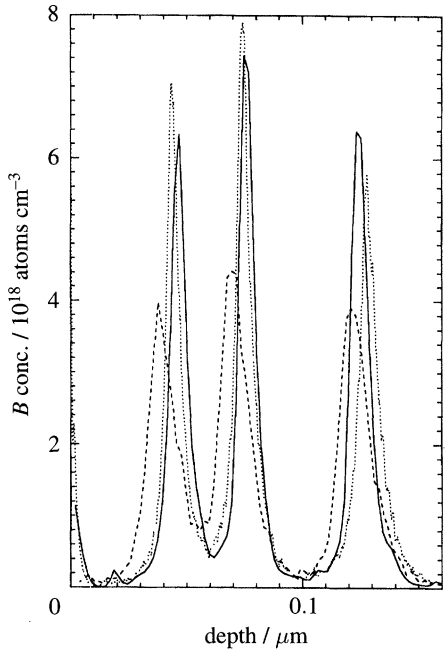

Figure 7. Measured SIMS data (9 keV O_2^+, normal incidence) and maximum entropy reconstruction for a set of 3 boron delta layers in silicon (dashed and solid lines respectively), compared with low energy (1.5 keV O_2^+, normal incidence) SIMS measurement without deconvolution (dotted line). The poor registration is partly caused by underdefinition, even though the data spacing was *ca.* 2 nm.

profile. Where the layers appear depends on the direction taken by the search as it approaches convergence.

The term 'depth resolution' ultimately refers to the degree to which features at

Downloaded from rsta.royalsocietypublishing.org

different depths can be distinguished in a depth profile. In SIMS, as in all areas of measurement, resolution parameters are a matter of definition, usually based on pragmatic considerations which can be formulated into a simple mathematical model. (Note that the concept of a single parameter representing the depth resolution implies linearity in the process.) In depth profiling at present, there is no defined term which is directly related to 'distinguishability' (e.g. a criterion like Rayleigh's in optics). The parameters used as figures of merit for depth resolution are derived from the slopes and widths of profiles of single features such as delta-layers or sharp concentration steps (Dowsett 1996). These tend to be smaller than the minimum separation at which two adjacent layers can be distinguished and are therefore optimistic. It is clear that the achievement of a particular depth resolution, however defined, is dependent on adequate sampling rate. One must take 10–30 samples per depth resolution interval to legitimately claim that depth resolution. Ultimately, this may mean taking several data points per atomic plane eroded, but figure 3 and equation (5.1) show that this is what is required for an adequate independent sample.

6. The noise model

Equation (4.1) implies a knowledge of σ_i. This can only be obtained from a model for the noise appropriate to the measuring system. The noise model determines whether χ^2 is accurately representative of the data. It is obvious that equation (4.4) cannot be satisfied if the σ_i are so small that χ^2 can never be less than or equal to N, for example. In a detailed study of SIMS depth profile data from a well maintained instrument (Chu & Dowsett 1996) we have only found evidence for a contribution to n from statistical fluctuations typical of a counting process, i.e. we find the noise to be Poissonian in form, with $\sigma_i = \sqrt{\bar{Y}_i}$, where $\bar{Y}_i$ is the mean. Then the probability P of a particular Y_i is

$$P(Y_i) = \frac{\bar{Y}_i^{Y_i}}{Y_i!} \mathrm{e}^{-\bar{Y}_i}. \tag{6.1}$$

The results are shown in figure 8. Earlier assumptions of a more complex form (Allen 1994) turn out be due to trying to describe noise measurements as partially Gaussian.

It is important for the success of forward methods that the noise model be assessed from experiment, and not assumed *a priori*.

The constrained maximization of the entropy (4.2) does not enforce any local correlation on the recovered data—for a given $\{C_{\mathrm{X}_i}\}$ the entropy is the same for any arrangement of the ordinates (see figure 5). The standard deviation in Poisson noise is a factor $\sqrt{2}$ larger than for than for the Gaussian case (although the *form* of the distribution tends to Gaussian for large Y), and for small Y (less than 200 counts) MaxEnt can legitimately reconstruct significant intensity into isolated ordinates. Skilling (1994) recommends pre-blurring of such data. An alternative would be to apply extra constraints on C_{est} based on prior knowledge. For example, if production details of the sample are available, the gradient of C_{est} might be restricted on diffusion grounds. If the sampling density were high (several ordinates per atomic plane spacing) then adjacent sets of ordinates could be constrained to have the same value (effectively reduced to one point). If there exists no *a priori* information to indicate that a low intensity feature is not due to an atomically sharp concentration step, pre blurring should be avoided in our view. It would be better improve the experiment or to seek a regularization parameter more suited to high dynamic range data.

Downloaded from rsta.royalsocietypublishing.org

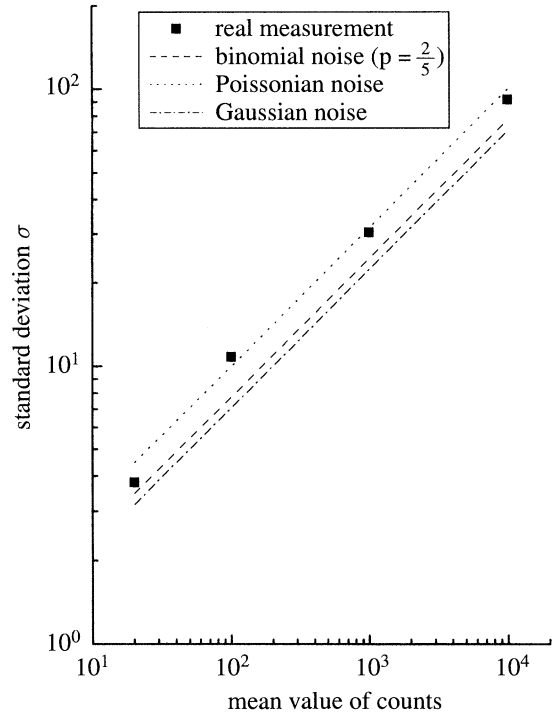

Figure 8. Measurement of noise amplitude on pulse counted SIMS depth profile data, compared the expected noise according to different models. The noise is Poissonian.

7. The response function

The response function for impurity X in matrix A can be measured, provided a sample is available containing a feature whose true width is significantly less than the depth resolution achievable. The linearity criteria in § 3 must be satisfied, both in respect of the reference material providing the response and the sample to be analysed. Elsewhere, both we and others describe the necessary samples and measurements in some detail (Dowsett *et al.* 1994*a*; Dowsett & Barlow 1993; Clegg & Gale 1991). Here, we comment on the problem of sampling, noise and finite range in the response itself. As measured, the response shares all the characteristics of the data described in § 2. Since SIMS measurements extend over at least 3 or 4 orders of magnitude in signal intensity, a uniform background (still Poissonian) of a few counts^{-1} is also typical. This background is not part of the response itself and needs to be removed before the data can be used in any deconvolution. It is also likely that the sampling interval $\Delta\phi_{\mathrm{resp}}$, and the range N_{resp} over which the response is valid, will differ from that of the data to be processed. These will therefore need to be modified by interpolation and extrapolation—some subjective judgements are inevitable. Perhaps the best way round the problem is to fit the measured response with a function (if one can be found) which replicates its shape as faithfully as possible. We use the convolution of two semi-infinite exponentials $\exp(\phi/\lambda_g)$ and $p(-\phi/\lambda_d)$ meeting at a cusp (physically unrealistic on their own) with a Gaussian of standard deviation σ (Dowsett *et al.* 1994*b*):

$$R_{\mathrm{X}} = (\phi) = k[(1 - \operatorname{erf} \xi_g)\mathrm{e}^{\zeta_g} + (1 + \operatorname{erf} \xi_d)\mathrm{e}^{-\zeta_d}], \tag{7.1}$$

Downloaded from rsta.royalsocietypublishing.org

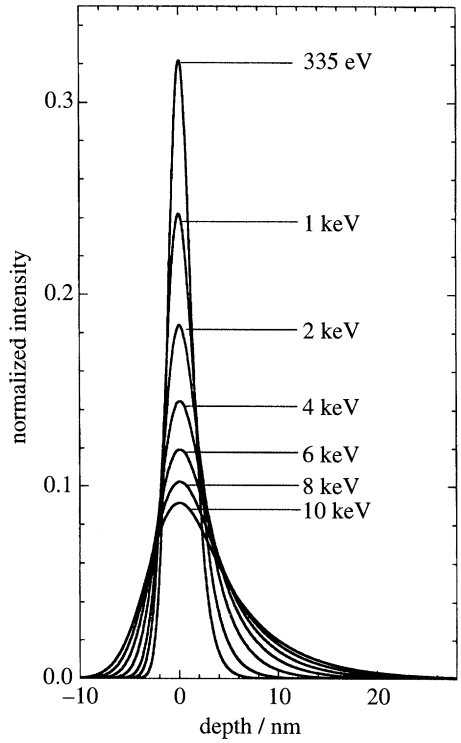

Figure 9. Equation (7.1) plotted using values of σ, λ_g and λ_d extracted from measured responses for a boron delta layer in silicon. (O_2^+ ions, normal incidence) The parameter is beam energy. On this scale, the differences between the measurement and the analytic expression are not discernible.

where

$$\xi_g = \frac{1}{\sqrt{2}}\left(\frac{\phi}{\sigma} + \frac{\sigma}{\lambda_g}\right), \quad \xi_d = \frac{1}{\sqrt{2}}\left(\frac{\phi}{\sigma} - \frac{\sigma}{\lambda_d}\right),$$

$$\zeta_g = \left(\frac{\phi}{\lambda_g} + \frac{\sigma^2}{2\lambda_g^2}\right), \quad \zeta_d = \left(\frac{\phi}{\lambda_d} - \frac{\sigma^2}{2\lambda_d^2}\right).$$

The parameter k normalizes the function. Figure 9 shows a set of such responses for data measured from a low temperature boron delta layer in silicon using normally incident O_2^+ ions. Other candidate functions have been suggested by Zalm & de Kruif (1993) and Hoffman (1992), for example. Equation (7.1) can be used to generate a response over any range of ϕ at a $\Delta\phi$ matching the experiment for use in forward or inverse numerical processes. It also has an analytic Fourier transform, which is useful under some circumstances for greatly speeding up convolution and deconvolution.

8. Ultra-high depth resolution profiling

Early predictions suggested that the depth resolution would peak for ion energies in the low keV region as a balance was struck between decreasing erosion rates (and therefore increased ion dose in the near-surface volume) and decreasing damage per ion (Andersen 1979). Recent experiments (e.g. figure 9) have shown that, for O_2^+ and inert gas ions, the depth resolution measured by various parameters continues

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Downloaded from rsta.royalsocietypublishing.org

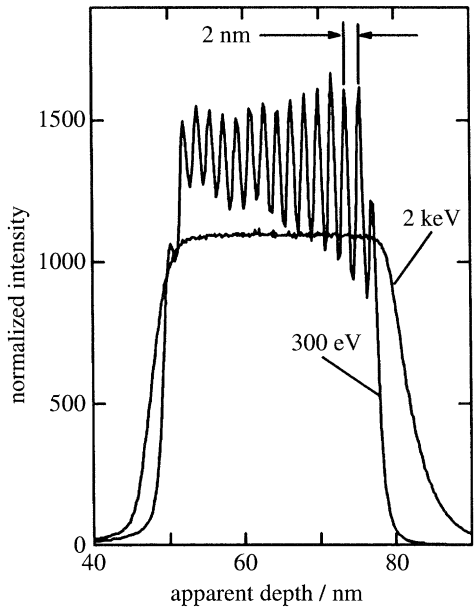

Figure 10. SIMS profiles of an alternating layer structure of silicon and germanium, with each layer 1 nm wide(O_2^+ ions, normal incidence). The 300 eV profile has a depth resolution of *ca.* 1 nm. The 2 keV profile shows an apparently homogeneous thick layer. (The small peak to the shallow side of this profile is an artifact due to Ge segregation in the 2 keV beam.) The two profiles are normalized to equal areas for comparison.

to improve down to energies of less than 250 eV (Clegg *et al.* 1996; Menyhard *et al.* 1995). However, producing a beam with sufficient current (greater than 100 nA) into a spot size of a few tens of μm suitable for SIMS depth profiling at these low energies has required the development of a new generation of floating ion columns where the beam transport is at high energy and the ions are retarded and focused onto the sample by retarding immersion lenses. With such a column it is possible to achieve depth resolutions of *ca.* 1 nm without resorting to deconvolution, and in some cases to measure the real chemical distributions of layers previously described as deltas.. Figure 10 shows a SIMS depth profile taken with 300 eV normally incident O_2^+ ions from a structure consisting of alternating 1 nm wide silicon and germanium layers grown by MBE. The deepening valley as a function of depth is intrinsic to the sample and is due to the stress induced buckling of the near surface layers. This type of sample is not suitable for deconvolution as it breaks the rules in § 2. In any case, as shown, SIMS profiles at greater than or equal to 1.5 keV indicate a homogeneous layer with no internal structure—no information for a deconvolution to recover.

9. Conclusions

Deconvolution is a necessary tool in the processing of SIMS data insofar as the measurement process remains linear. It is no replacement for improved experimental technique or better instrumentation and it should not be regarded as simply a means of 'sharpening up' poor data. The response in SIMS depth profiling is significantly more destructive of information than a Gaussian of similar full width at half maximum—as equation (7.1) implies, the decay is by no means as rapid. A forward method based on MaxEnt can recover a C_{est} which *could* have given rise to

Downloaded from rsta.royalsocietypublishing.org

the measured data according to a statistical criterion because the recovered data are not limited by smoothing (e.g. Cooke *et al.* 1996*a*, *b*). An FFT-based method cannot recover data which could have given rise to the measurement—although it may give a useful estimate (e.g. Herzel *et al.* 1995).

In other fields, where deconvolution is used more routinely, the objective is often to recover unknown amplitudes for known peak shapes at predictable abscissa, i.e. in spectroscopy, or to recover a two-dimensional intensity distribution containing recognizable objects, e.g. sharpening up a blurred photograph. There are then simple ways of rejecting spurious data produced by over ambitious deconvolution. The closest sputter profiling comes to this is in examples like figure 6—however, neither the abscissa (depth) nor the actual chemical distribution in the delta (peak shape) are known in these cases. Only very accurate profiling at sub-keV beam energies can provide a check on the deconvolved concentration.

It is clear that with ultra low energy SIMS data whose intrinsic resolution is *ca.* 1 nm, deconvolution is not always necessary (and if it were to be used one would have the ultimate depth resolution possible—1 atomic plane or so). However, if depth and lateral resolution need to be combined (for example, in lateral dopant profiling (Cooke *et al.* 1996*c*), it will never be possible to operate at the lowest attainable energies (probe diameters and current densities are compromized)). Continued study of the statistical properties of SIMS and other sputter profile data, and the testing and development of deconvolution methods combined with different approaches to the depth profiling (e.g. Hsu *et al.* 1995), can still further enhance the technique.

Bob Collins died while the invited speakers at the Discussion Meeting in June 1995 were being selected—otherwise he would have made the presentation. I am indebted to Anne Collins for permission to include his name on the paper, and to use material from notes he sent me during our collaboration on MaxEnt methods. I hope he would have been pleased with the result.

Work included here was funded by the U.K. Silicon Towards 2000 Programme (Projects IED 1559 and 1540), the R. W. Paul Instrument Fund and EPSRC (notably GR/H65177 and GR/K32715)

References

Allen, P. N. 1994 The quantification of SIMS depth profiles by maximum entropy reconstruction. Ph.D. thesis, University of Warwick.

Allen, P. N., Dowsett, M. G. & Collins, R. 1993 *Surf. Interface Analysis* **20**, 696–702.

Andersen, H. H. 1979 *Appl. Phys.* **18**. 131–140.

Boudewijn, P. R., Ackerboom, H. W. & Kempeners, M. N. C. 1984 *Spectrochim. Acta.* B **39**. 1567–1571.

Briggs, D. 1992 In *Practical surface analysis* (ed. D. Briggs & M. P. Seah), vol. II, ch. 7. Chichester: Wiley.

Chu, D. & Dowsett, M. G. 1996 *J. Appl. Phys.* (In the press.)

Clegg, J. B. & Gale, I. G. 1991 *Surf. Interface Analysis* **17**. 190–196.

Clegg, J. B., Smith, N. S., Dowsett, M. G., Theunissen, M. J. J. & de Boer, W. B. 1996 *J. Vac. Sci. Technol.* A **14**, 2645–2650.

Collins, R. 1986 *Radiat. Eff.* **98**, 1–20.

Collins, R., Jimenez-Rodriguez, J. J., Wadsworth, M. & Badheka, R. 1988 *J. Appl. Phys.* **64**, 1120–1124.

Cooke, G. A., Dowsett, M. G., Allen, P. N., Collins, R. & Miethe, K. 1996*a* *J. Vac. Sci. Technol* B **14**, 132–135.

Cooke, G. A., Dowsett, M. G. & Phillips, P. J. 1996*b* *J. Vac. Sci. Technol.* B **14**, 283–286.

Downloaded from rsta.royalsocietypublishing.org

Cooke, G. A., Pearson, P., Gibbons, R., Dowsett, M. G. & Hill, C. 1996*c* *J. Vac. Sci. Technol.* B **14**, 348–352.

Dowsett, M. G. 1996 In *Secondary ion mass spectrometry SIMS X* (ed. A. Benninghoven *et al.*). Chichester: Wiley. (In the press.)

Dowsett, M. G. & Barlow, R. D. 1994 *Analysis Chim. Acta* **297**, 253–275.

Dowsett, M. G., Barlow, R. D. & Allen, P. N. 1994 *J. Vac. Sci. Technol.* B **12**, 186–198.

Gull, S. F. 1994 In *Maximum entropy in action* (ed. B. Buck & V. A. Macaulay), ch. 7. Oxford: Clarendon.

Herzel, F., Ehwald, K. E., Hienemann, B., Krüger, D., Kurps, R., Röpke, W. & Zeindl, H. P. 1995 *Surf. Interface Analysis* **23**, 764–770.

Hofman, S. 1992 *J. Vac. Sci. Technol.* B **10**, 316–322.

Hsu, C. M., Sharma, V. K. M., Ashwin, M. J. & McPhail, D. S. 1995 *Surf. Interface Analysis* **23**, 665–672.

Levi Seti, R., Hallégot, P., Girod, C., Chabala, J. M., Li, J., Sodonis, A. & Wolbach, W. 1991 *Surf. Sci.* **246**, 94–106.

Littmark, U. & Hofer, W. O. 1980 *Nucl. Instrum. Meth.* **168**, 329–342.

Menyhard, M., Barna, A., Biersack, J. P., Järrendahl, K. & Sundgren, J. E. 1995 *J. Vac. Sci. Technol.* A **13**, 1999–2004.

Myrheim, J. & Rue, H. 1992 *Graphical Models and Image Processing* **54**, 223–238.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992*a* In *Numerical recipies in C*, ch. 18, § 4. Cambridge University Press.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992*b* In *Numerical recipies in C*, ch. 13, § 1. Cambridge University Press.

Rüdenauer, F. G. 1994 *Analysis Chim. Acta* **297**, 197–230.

Skilling, J. 1994 In *Maximum entropy in action* (ed. B. Buck & V. A. Macaulay), ch. 2. Oxford: Clarendon.

Skilling, J. & Gull, S. F. 1985 Algorithms and applications. In *Maximum-entropy and Bayesian methods in inverse problems* (ed. C. Ray Smith & W. T. Grandy Jr), pp. 83–131. Dordrecht: Reidel.

Wang, X. S., Pechman, R. J. & Weaver, J. H. 1995 *J. Vac. Sci. Technol.* B **13**, 2031–2040.

Willaims, P. & Baker, J. E. 1980 *Appl. Phys. Lett.* **36**, 842–845.

Wittmaack, K. 1980 *Nucl. Instrum. Meth.* **168**, 343–356.

Zalm, P. C. & de Kruif, R. C. M. 1993 *Appl. Surface Sci.* **70/71**, 73–78.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES